CONTINUOUS TIME THRESHOLD AUTOREGRESSIVE MODELS

By Robin John Hyndman

A thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy
in the Department of Statistics
at the University of Melbourne.

December 1992

Abstract

This thesis considers continuous time autoregressive processes defined by stochastic differential equations and develops some methods for modelling time series data by such processes.

The first part of the thesis looks at continuous time *linear* autoregressive (CAR) processes defined by *linear* stochastic differential equations. These processes are well-understood and there is a large body of literature devoted to their study. I summarise some of the relevant material and develop some further results. In particular, I propose a new and very fast method of estimation using an approach analogous to the Yule–Walker estimates for discrete time autoregressive processes. The models so estimated may be used for preliminary analysis of the appropriate model structure and as a starting point for maximum likelihood estimation.

A natural extension of CAR processes is the class of continuous time *threshold* autoregressive (CTAR) processes defined by *piecewise linear* stochastic differential equations. Very little attention has been given to these processes with a few isolated papers in the engineering and probability literature over the past 30 years and some recent publications by Tong and Yeung (summarised in Tong, 1990). I consider the order one case in detail and derive conditions for stationarity, equations for the stationary density, equations for the first two moments and discuss various approximating Markov chains. Higher order processes are also discussed and several approaches to estimation and forecasting are developed.

Both CAR and CTAR models are fitted to several real data sets and the results

are compared. A rule-based procedure is suggested for fitting these models to a given time series.

One difficulty with using non-linear models (such as CTAR models) is that the forecast densities are not Gaussian, not symmetric and often multi-modal. Therefore, it is inappropriate to just consider the mean and standard deviation of forecasts. It is argued that for non-Gaussian forecast densities, highest density regions should be used when describing forecasts. An algorithm which enables the rapid construction of highest density regions given a probability density function is developed. The methods described are applicable to univariate data with a continuous density containing any number of local modes. The density may be known analytically or may be estimated from observations or simulations. Code for the algorithm is provided in the C language.

Declaration

I hereby declare that:

1 This thesis is less than 100,000 words in length.

2 This thesis contains no material which has been accepted for the award of any other degree or diploma.

3 To the best of my knowledge, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Acknowledgments

I am greatly indebted to my supervisor, Dr Gary Grunwald for his time, patience, enthusiasm and support in assisting me during my candidature. I would also like to thank Professor Peter Brockwell for many useful discussions in connection with this material and for his generous hospitality during my visit to Colorado State University in February 1991. Without both Gary and Peter, this work would not have been done.

I would also like to thank the Federal Government for supporting this work with an Australian Postgraduate Award, BHP Melbourne Research Laboratories for supporting this work with a BHP Research Grant, and Dr Julie Sims of BHP MRL for advice on the application of this material to BHP data. I am grateful to Julie and Vladimir Panjkovic for supplying data relating to BHP steel production.

Others who deserve thanks are Dr Fima Klebaner and Dr Kais Hamza of the University of Melbourne and Dr Claes Jogreus of Colorado State University for their ready suggestions on Chapter 4, Professor Jan de Gooijer of the University of Amsterdam for some valuable comments on Chapter 6 and several anonymous referees for comments on those parts of this work which have appeared as journal articles. In addition, Rosemary White for secretarial assistance and humorous relief and my parents who supported me during much of my education and provided much encouragement (and cups of tea). Finally, special thanks to my wife Leanne who has given me steady support in countless ways throughout my work.

CONTENTS

Preface

The material in Sections 2.3, 2.4 and 3.1 of this thesis has been previously published as the journal article Hyndman (1992).

Most of the material in Chapters 4–6 has appeared in two journal articles: Brockwell, Hyndman and Grunwald (1991) and Brockwell and Hyndman (1992). These were preliminary reports of the results that are presented in this thesis.

Chapter 7 has been submitted to the *Journal of Graphical and Computational Statistics* and is currently being refereed for publication.

Chapter 1
INTRODUCTION

Traditionally, time series analysis has been concerned with discrete time linear models. However, the twin assumptions of discrete time and linearity often limit the effectiveness of the model in describing the underlying process.

The strait-jacket of linearity prevents the models capturing such commonly occurring phenomena as asymmetric but periodic behaviour, time-irreversibility, non-Gaussian marginal distributions and the occurrence of occasional bursts of outlying observations. Linear models are popular because they lead to elegant mathematics and fast algorithms, but they are not realistic descriptions of the world. Real data are frequently non-linear.

The importance of non-linear models in time series analysis has been recognised increasingly over the past ten years. A number of discrete time non-linear processes have been introduced and found valuable for the modelling of observed series. Among these processes are the discrete time threshold models discussed extensively in the books of Howell Tong (1983, 1990). It was Tong's threshold models that motivated the idea of continuous time threshold autoregressive

models developed in this thesis.

The restriction to a discrete time perspective often masks the inherent continuity of the underlying process. For example, many physical processes are inherently continuous and observed at discrete time points. Economic data are often the result of large numbers of microeconomic decisions taken at many different times and so may be considered as continuous time series. Frequently data are observed at irregularly spaced intervals; time series with missing values may be considered in this context. In each of these cases, the underlying process from which observations are taken may be considered a function in continuous time.

This thesis considers time series models in which the restrictions to discrete time and linearity are dropped.

1.1 Discrete Time Threshold Models

Before describing a continuous time threshold autoregressive model, it is appropriate to briefly mention the discrete time threshold model which led to the development of the models in this thesis.

A discrete-time self-exciting threshold autoregressive (SETAR) process $\{X_t\}$ with delay parameter d is a solution of the equations

$$X_t = \sum_{j=1}^{p} a_{i,j} X_{t-j} + b_i + \sigma_i e_t, \qquad r_{i-1} \leq X_{t-d} < r_i, \tag{1.1.1}$$

where $t = 0, \pm 1, \pm 2, \cdots$ and $i = 1, \cdots, l$. The delay parameter, d, is a positive integer and $\{e_t\}$ is a white noise sequence with unit variance.

Thus, the real line is partitioned into l regimes and X_t satisfies one of l autoregressive equations depending on the regime in which X_{t-d} falls. When $l = 1$, $\{X_t\}$ is a linear autoregressive process.

The autoregressive order p may be allowed to differ between regimes by simply letting p_i be the order in regime i and setting $a_{i,j} = 0$ for all $j > p_i$. Then $p = \max_i p_i$.

This non-linear time series model was proposed by Tong (1978) and is discussed in detail by Tong (1990). It is capable of reproducing features which are frequently observed in naturally occurring time series but which cannot be captured with linear models.

Owing to the difficulty of analysing non-linear difference equations, theoretical results for the general SETAR process are far from complete. However, necessary and sufficient conditions for existence of a stationary solution of (1.1.1) have been derived by Chan et al. (1985) for the case $p = 1$ and $d = 1$. Similar conditions for the case $p = 1$ and $l = 2$ are given by Chen and Tsay (1991).

Tests for the appropriateness or otherwise of linear models are available; for a survey and comparison of current methods, see Petruccelli (1990). Identification and estimation has also been the subject of several papers; for example, Tsay (1989) and Davies, Pemberton and Petruccelli (1988). The computer package STAR of Tong (1990) can be used for testing and fitting discrete time threshold models with up to three thresholds.

1.2 Continuous Time Threshold Models

There are several good reasons for investigating continuous time analogues of (1.1.1). For a continuously evolving process observed at discrete time intervals it is natural to allow the possibility of a threshold crossing at *any* time. Furthermore, the process should be allowed to switch between regimes any number of times between successive observations. It would also be useful to allow a delay parameter, d, which is not necessarily a positive integer multiple of the observation spacing. The case $d = 0$ seems to be of particular importance since for many naturally occurring continuous processes, their future depends only the *current* state of the process. Yet discrete time threshold models require $d > 0$. In this thesis, I shall restrict myself to this case of zero delay.

Of course, the advantages generally associated with continuous time models are also associated with continuous time threshold autoregressive models. That is, they may be used to model irregularly spaced data and are particularly appropriate for modelling time series based on an underlying process that is inherently continuous.

The CTAR processes described here are analogous to the SETAR processes with the stochastic piecewise linear difference equations replaced by stochastic piecewise linear differential equations. Thus, they may be written as

$$X^{(p)}(t) + \alpha_{p-1,i}X^{(p-1)}(t) + \cdots + \alpha_{0,i}X(t) + \beta_i = \sigma_i Z(t), \quad r_{i-1} < X(t) < r_i,$$

where $-\infty < t < \infty$, $i = 1, \ldots, l$, $-\infty = r_0 < r_1 < \cdots < r_l = \infty$, $\{X(t)\}$ denotes the time series, $X^{(j)}(t)$ denotes the jth derivative of $X(t)$ and $\{Z(t)\}$ denotes continuous time Gaussian white noise.

As we shall see, the analysis of continuous time threshold autoregressive models can be reduced to the study of diffusion processes with piecewise linear coefficients. In this context, CTAR processes were the subject of a few isolated papers in the engineering and probability literature during the 1960s, namely Khazen (1961), Atkinson (1967) and Atkinson and Caughey (1968).

CTAR processes are special cases of more general stochastic differential equations and diffusion processes. Several authors have considered general diffusion processes with discontinuous drift coefficient; for example, Stroock and Varadhan (1969), Zvonkin (1974), Conway (1971, 1972) and Lebedev (1976, 1978). However, the results developed by these authors require a continuous dispersion coefficient and so they are not generally applicable to CTAR processes. Diffusions with discontinuous dispersion coefficient do not seem to be well-understood.

In the time series literature, Yeung (1989), Tong (1990) and Tong and Yeung (1990, 1991) have considered CTAR processes. They have developed tests of threshold-type non-linearity and have suggested a method of estimation. I shall compare their estimation method with the one developed in Chapter 5.

1.3 Summary of Contents

Continuous time linear autoregressive (CAR) processes are discussed in Chapter 2. In this case, there is just one regime ($l = 1$) and so the parameters are constant. There is a large body of literature concerning such processes and most of this chapter is a summary of well-known results. However, the emphasis on the derivative covariance function in Sections 2.3 and 2.4 is a different

perspective from that usually presented and leads to the development of a new method of estimation in the following chapter.

In Chapter 3, I consider the problem of estimating CAR models from a time series of observations. In particular, I propose a new and very fast method of estimation using an approach analogous to the Yule–Walker estimates for discrete time autoregressive processes. The models so estimated may be used for preliminary analysis of the appropriate model structure and as a starting point for maximum likelihood estimation (Jones, 1981). A modelling strategy incorporating these ideas is outlined at the end of Chapter 3.

In Chapter 4, I consider the CTAR process with $p = 1$ in detail and derive conditions for stationarity. It is shown that the stationary distributions, moments and conditional expectations can be computed from the forward and backward Kolmogorov equations with appropriate boundary conditions. It turns out that the boundary conditions at the thresholds are crucial to the definition and properties of the process. Three different boundary conditions are considered and the properties of the associated processes are discussed. A useful approach to CTAR processes is via approximating discrete time Markov chains. It is shown that different Markov chains implicitly determine different boundary conditions.

Higher order CTAR processes are discussed in Chapter 5, although in considerably less detail than processes of order one. Difficulties with rigorously defining the process are discussed and it is shown how the conditional moments of the forecast density can be computed.

Chapter 6 develops several approaches to estimation and forecasting. These

are compared to the approach developed by Tong and Yeung. Several simulated and real data sets are used to compare the estimation methods and to compare the fitted CTAR models with linear CAR models. The modelling strategy outlined at the end of Chapter 3 for fitting CAR models is extended to incorporate CTAR models.

One difficulty with forecasting CTAR models is that the forecast densities are not Gaussian, not symmetric and may be multi-modal. Therefore, it is inappropriate to just consider the mean and standard deviation of forecasts. In Chapter 6, it is argued that for non-Gaussian forecast densities, it is appropriate to use highest density regions when describing forecasts. An algorithm which enables the rapid construction of highest density regions given a probability density function is described in Chapter 7. The methods described are applicable to univariate data with a continuous density containing any number of local modes. The density may be known analytically or may be estimated from observations or simulations. A C version of the algorithm is also provided.

Chapter 2

CONTINUOUS TIME LINEAR AUTOREGRESSIVE PROCESSES

In this chapter, I investigate continuous time linear autoregressive (CAR) processes and develop the requisite theory that will be used in Chapter 3 for deriving methods for estimating these models.

The definition of a CAR process is discussed in Section 2.1 and the connection between these and diffusion processes is emphasised. This connection is important in the development of boundary conditions for threshold models. Conditions for stationarity are discussed in Section 2.2 and the results from this discussion will be used in investigating the stationarity of estimators introduced in later chapters.

Sections 2.3 and 2.4 present a somewhat different perspective from that normally presented in continuous time series analysis. The autocovariance and derivative covariance functions are introduced in Section 2.3 and some of their properties are developed. Yule–Walker type equations are derived for CAR models in Section 2.4. These will be fundamental in developing new estima-

tors for continuous time autoregressive models.

Finally, Section 2.5 discusses the nature of the process formed by observations of a CAR(p) process taken at regular intervals.

2.1 Definition of a CAR Process

A continuous time autoregressive process of order p and mean $-\beta/\alpha_0$ may be represented by the equation

$$X^{(p)}(t)+\alpha_{p-1}X^{(p-1)}(t)+\cdots+\alpha_0 X(t)+\beta = \sigma Z(t), \qquad -\infty < t < \infty, \quad (2.1.1)$$

where $\{Z(t)\}$ denotes continuous time Gaussian white noise. Of course, $\{Z(t)\}$ does not exist as a stochastic process in the usual sense and so this is not a rigorous definition of a CAR process. Nevertheless, it is a useful description and is intuitively appealing. A rigorous definition of a CAR process is given below.

2.1.1 CAR(1) processes

Define a linear, continuous time autoregressive process of order one (denoted by CAR(1)) to be a stationary solution of the stochastic differential equation

$$dX(t) + \alpha_0 X(t)dt + \beta dt = \sigma dW(t), \qquad t \geq 0, \qquad (2.1.2)$$

where $\{W(t)\}$ denotes standard Brownian motion, $X(0) \stackrel{d}{=} N\left(-\frac{\beta}{\alpha_0}, \frac{\sigma^2}{2\alpha_0}\right)$, $\alpha_0 > 0$ and $\sigma > 0$. Under these conditions, this equation has the unique stationary

solution

$$X(t) = e^{-\alpha_0 t}X(0) - \frac{\beta}{\alpha_0}\left[1 - e^{-\alpha_0 t}\right] + \sigma\int_0^t e^{-\alpha_0(t-y)}dW(y).$$

Equivalently, a CAR(1) process can be defined as a stationary diffusion process with drift coefficient $\mu(x) = -\alpha_0 x - \beta$ and diffusion coefficient $\Sigma(x) = \sigma^2/2$. If $\beta = 0$, this is a stationary Ornstein–Uhlenbeck process. The generator of this diffusion process is defined for functions on $\mathbb{R}$ with bounded continuous second derivatives by

$$\mathcal{G}f(x) = \mu(x)f'(x) + \Sigma(x)f''(x).$$

For inference, we shall use the functions

$$\begin{aligned} m(x,t) &:= E(X(t)|X(0) = x) = e^{-\alpha_0 t}x - \frac{\beta}{\alpha_0}\left[1 - e^{-\alpha_0 t}\right], \\ m_2(x,t) &:= E(X^2(t)|X(0) = x) \\ \text{and} \quad v(x,t) &:= m_2(x,t) - m^2(x,t) = \sigma^2\int_0^t e^{-2\alpha_0(t-u)}du = \frac{\sigma^2}{2\alpha_0}\left[1 - e^{-2\alpha_0 t}\right]. \end{aligned}$$

2.1.2 CAR(p) processes

In order to define a CAR(p) process rigorously, it is convenient to introduce the p-variate process $\boldsymbol{S}(t) := \left[X(t), X^{(1)}(t), \cdots, X^{(p-1)}(t)\right]^T$ satisfying the stochastic differential equation,

$$d\boldsymbol{S}(t) = A\boldsymbol{S}(t)dt - \beta\boldsymbol{b}dt + \sigma\boldsymbol{b}dW(t), \qquad t \geq 0, \tag{2.1.3}$$

where $\{W(t)\}$ denotes standard Brownian motion, $\sigma > 0$,

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -\alpha_0 & -\alpha_1 & \cdots & \cdots & -\alpha_{p-1} \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

and

$$\boldsymbol{S}(0) \stackrel{d}{=} N\left(-\frac{\beta}{\alpha_0}\boldsymbol{a}\ ,\ \sigma^2 \int_0^\infty e^{Au}\boldsymbol{b}\boldsymbol{b}^T e^{A^T u} du\right) \tag{2.1.4}$$

is independent of $\{W(s), s \geq 0\}$. Here $\boldsymbol{a} = [1, 0, \ldots, 0]^T$.

Then $\{X(t)\}$ is stationary if and only if all the eigenvalues of A have negative real parts (Arató, 1982, pp.118–119). Under this stationarity condition, $\{X(t)\}$ is defined to be a CAR(p) process.

Equation (2.1.1) should be regarded as a formal representation of the well-defined Itô differential equation (2.1.3).

The solution of (2.1.3) is the Itô integral

$$\boldsymbol{S}(t) = e^{At}\boldsymbol{S}(0) + \sigma \int_0^t e^{A(t-y)}\boldsymbol{b} dW(y) - \beta \int_0^t e^{A(t-y)}\boldsymbol{b} dy. \tag{2.1.5}$$

The CAR(p) process $\{X(t)\}$ can also be specified as the first component of the stationary multivariate diffusion process $\{\boldsymbol{S}(t)\}$ with drift coefficient vector $\boldsymbol{\mu}(\boldsymbol{x}) = A\boldsymbol{x} - \beta\boldsymbol{b}$ and diffusion coefficient matrix $\Sigma(\boldsymbol{x}) = \sigma^2\boldsymbol{b}\boldsymbol{b}^T/2$. The

generator of $\boldsymbol{S}(t)$ satisfies

$$\begin{aligned}
&\mathcal{G}f(x) \\
&= \Sigma(\boldsymbol{x})\frac{d^2 f}{d\boldsymbol{x}d\boldsymbol{x}^T} + \boldsymbol{\mu}^T(\boldsymbol{x})\frac{df}{d\boldsymbol{x}} \\
&= \frac{\sigma^2}{2}\frac{\partial^2 f}{\partial x_p^2} - \alpha_0 x_1 \frac{\partial f}{\partial x_p} - \cdots - \alpha_{p-1} x_p \frac{\partial f}{\partial x_p} - \beta\frac{\partial f}{\partial x_p} + x_p\frac{\partial f}{\partial x_{p-1}} + \cdots + x_2\frac{\partial f}{\partial x_1}
\end{aligned}$$

for all twice continuously differentiable functions f on $\mathbb{R}^p$ such that each term on the right is bounded.

I shall denote the mean-corrected process by $\boldsymbol{S}_*(t) = \boldsymbol{S}(t) + \frac{\beta}{\alpha_0}\boldsymbol{a}$. The components of $\boldsymbol{S}_*(t)$ shall be denoted by $X_*^{(j)}(t)$, $0 \le j \le p-1$. Note that $\boldsymbol{S}_*(t)$ satisfies the state space equations

$$\begin{aligned}
& d\boldsymbol{S}_*(t) = A\boldsymbol{S}(t)dt + \sigma\boldsymbol{b}dW(t), \qquad t \ge 0 \\
\text{and} \qquad & X_*(t) = \boldsymbol{a}^T\boldsymbol{S}_*(t).
\end{aligned}$$

As for CAR(1) processes, it will be useful to define the functions

$$m(\boldsymbol{x},t) := E(X(t)|\boldsymbol{S}(0)=\boldsymbol{x}) = \boldsymbol{a}^T\left[e^{At}\boldsymbol{x} - \beta\int_0^t e^{A(t-u)}\boldsymbol{b}du\right], \quad (2.1.6)$$

$$m_2(\boldsymbol{x},t) := E(X^2(t)|\boldsymbol{S}(0)=\boldsymbol{x}) \quad (2.1.7)$$

$$\text{and} \quad v(\boldsymbol{x},t) := m_2(\boldsymbol{x},t) - m^2(\boldsymbol{x},t) = \boldsymbol{a}^T\left[\sigma^2\int_0^t e^{A(t-u)}\boldsymbol{b}\boldsymbol{b}^T e^{A^T(t-u)}du\right]\boldsymbol{a}. \quad (2.1.8)$$

2.1.3 Existence of stochastic derivatives

Some general results for stochastic differential equations (see, for example, Jazwinski, 1970, p.111) may be applied to $\{X(t)\}$ to obtain the following

important results:

1. $\{X^{(j)}(t)\}$ is mean square continuous for $j = 0, 1, \cdots, p-1$ and mean square differentiable for $j = 0, 1, \cdots, p-2$;

2. $\{X^{(p-1)}(t)\}$ is not mean square differentiable;

3. $\{X^{(j)}(t)\}$ is equivalent to a process with almost all sample functions continuous for $j = 0, 1, \cdots, p-1$. Hence $\{X(t)\}$ is equivalent to a process with almost all sample functions possessing continuous derivatives up to order $p-1$.

4. $X^{(j)}(t)$ is normally distributed for $j = 0, 1, \cdots, p-1$ and for each t.

2.2 Conditions for Stationarity

The eigenvalues of A are simply the roots of the characteristic equation

$$\mathcal{C}(z) = \sum_{j=0}^{p} \alpha_j z^j = 0, \qquad \text{where } \alpha_p = 1.$$

Thus, for stationarity, we require that $\mathrm{Re}(\lambda) < 0$, for each root λ of $\mathcal{C}(z)$.

For $p = 1$ and $p = 2$ the characteristic roots have negative real parts if and only if all coefficients are positive. In fact, the positivity of the coefficients is a sufficient condition for stationarity for all p. This result follows from the next theorem which has been adapted from Kaplan (1962).

THEOREM 2.2.1 *Let $\phi_n(s) = p_0 + p_1 s + \cdots + p_n s^n$ where each p_i is real and all the roots of $\phi_n(s)$ have negative real parts. Then the coefficients p_i all have the same sign.*

PROOF: We shall prove the result by induction. Suppose $p_n > 0$. (If $p_n < 0$, simply multiply the polynomial by -1.) Then we need to show that $p_i > 0$, $i = 0, 1, \ldots, n-1$.

Now let $n = 1$ and let s^* be a root of $\phi_1(s)$. Then $\phi_1(s^*) = p_0 + p_1 s^* = 0$ so that $p_0 = -p_1 s^* > 0$ since $s^* < 0$ and $p_1 > 0$. So the theorem is true for $n = 1$. Now suppose the theorem is true for $n = k$.

If $\phi_{k+1}(s)$ has a real root s^*, then $\phi_{k+1}(s) = p_{k+1}(s - s^*)\xi(s)$ where $p_{k+1} > 0$ and $\xi(s)$ is a monic polynomial of degree k. Now $s^* < 0$ by assumption and $\xi(s)$ has roots with negative real parts. Hence, by induction, $\xi(s)$ has positive coefficients. Thus, $\phi_{k+1}(s)$ is a product of factors with positive coefficients and so, itself, has positive coefficients.

If $\phi_{k+1}(s)$ has no real root, it has a pair of conjugate complex roots $\alpha \pm \beta i$ with $\alpha < 0$. So

$$\begin{aligned} \phi_{k+1}(s) &= p_{k+1}(s - \alpha - \beta i)(s - \alpha + \beta i)\xi(s) \\ &= p_{k+1}(s^2 - 2\alpha s + \alpha^2 + \beta^2)\xi(s) \end{aligned}$$

where $\xi(s)$ is a monic polynomial of degree k. Since $\alpha < 0$, all factors have positive coefficients and so $\phi_{k+1}(s)$ has positive coefficients.

Thus, the theorem is true for $k + 1$. Hence, by induction, the theorem is true for all $n \geq 1$.

□

Other useful tests for stationarity are obtained via the Routh–Hurwitz and Liénard–Chipart criteria

THEOREM 2.2.2 *Let* $\phi_n(s) = p_0 + p_1 s + \cdots + p_n s^n$ *where each* p_i *is real and* $p_n > 0$*, and define the determinants*

$$\Delta_1 = p_{n-1} \qquad \Delta_2 = \begin{vmatrix} p_{n-1} & p_n \\ p_{n-3} & p_{n-2} \end{vmatrix} \qquad \Delta_3 = \begin{vmatrix} p_{n-1} & p_n & 0 \\ p_{n-3} & p_{n-2} & p_{n-1} \\ p_{n-5} & p_{n-4} & p_{n-3} \end{vmatrix}$$

$$\cdots \qquad \Delta_k = \begin{vmatrix} p_{n-1} & p_n & 0 & 0 & 0 & \cdots & 0 \\ p_{n-3} & p_{n-2} & p_{n-1} & p_n & 0 & \cdots & 0 \\ \vdots & \vdots & & & & & \vdots \\ p_{2k-1} & p_{2k-2} & \cdots & \cdots & \cdots & \cdots & p_k \end{vmatrix} \qquad \cdots$$

where $k \leq p$ *and* $p_i = 0$ *for* $i < 0$ *and* $i > n$*. Then each of the following sets of conditions is necessary and sufficient for all the roots of* $\phi_n(s)$ *to lie in the half plane* $Re(s) < 0$*.*

1. $\Delta_1, \Delta_2, \cdots, \Delta_n$ *are all positive;*

2. $p_n > 0,\ p_0 > 0,\ p_2 > 0,\ \ldots;\quad \Delta_1 > 0,\ \Delta_3 > 0,\ \ldots;$

3. $p_n > 0,\ p_0 > 0,\ p_2 > 0,\ \ldots;\quad \Delta_2 > 0,\ \Delta_4 > 0,\ \ldots;$

4. $p_n > 0,\ p_0 > 0,\ p_1 > 0,\ p_3 > 0,\ \ldots;\quad \Delta_1 > 0,\ \Delta_3 > 0,\ \ldots;$

5. $p_n > 0,\ p_0 > 0,\ p_1 > 0,\ p_3 > 0,\ \ldots;\quad \Delta_2 > 0,\ \Delta_4 > 0,\ \ldots.$

The first of these is due to Routh (1877, 1905) and Hurwitz (1895), the others to Liénard and Chipart (1914). For discussion and proofs of these results, see Kaplan (1962, pp.408–413) and Gantmacher (1960, pp221f).

Applying the preceding results to linear CAR(p) processes yields a variety of statements about stationarity that involve conditions on the model coefficients.

COROLLARY 2.2.3 *Let $\boldsymbol{S}(t)$ satisfy (2.1.3) and (2.1.4), and let $X(t)$ be the first component of $\boldsymbol{S}(t)$. Then*

1. *$X(t)$ is a stationary process if each $\alpha_j > 0$, $j = 0, 1, \cdots, p-1$;*

2. *$X(t)$ is a stationary process if and only if one of the following is true*

a) $\Delta_j > 0$, $j = 1, 2, \ldots, p$;

b) $\alpha_0 > 0$, $\alpha_2 > 0$, $\alpha_4 > 0$, $\ldots$; $\Delta_1 > 0$, $\Delta_3 > 0$, $\ldots$;

c) $\alpha_0 > 0$, $\alpha_2 > 0$, $\alpha_4 > 0$, $\ldots$; $\Delta_2 > 0$, $\Delta_4 > 0$, $\ldots$;

d) $\alpha_0 > 0$, $\alpha_1 > 0$, $\alpha_3 > 0$, $\ldots$; $\Delta_1 > 0$, $\Delta_3 > 0$, $\ldots$;

e) $\alpha_0 > 0$, $\alpha_1 > 0$, $\alpha_3 > 0$, $\ldots$; $\Delta_2 > 0$, $\Delta_4 > 0$, $\ldots$.

COROLLARY 2.2.4 *For the cases $1 \leq p \leq 4$, the following conditions are necessary and sufficient for a stationary solution of (2.1.3) and (2.1.4) to exist.*

$p = 1$ $\alpha_0 > 0$;

$p = 2$ $\alpha_0 > 0$ *and* $\alpha_1 > 0$;

$p = 3$ $\alpha_0 > 0$, $\alpha_2 > 0$ *and* $\alpha_1\alpha_2 - \alpha_0 > 0$;

$p = 4$ $\alpha_0 > 0$, $\alpha_1 > 0$, $\alpha_3 > 0$, $\alpha_1\alpha_2\alpha_3 - \alpha_0\alpha_3^2 - \alpha_1^2 > 0$.

2.3 Covariance Functions

2.3.1 The autocovariance function

Let $\gamma(h) = \text{Cov}\,[X(t+h), X(t)]$ denote the autocovariance function (ACVF) of the CAR(p) process $\{X(t)\}$ and let $f(\omega)$ denote the corresponding spectral

density function. Then, following Doob (1953, p.543), we obtain

$$f(\omega) = \frac{\sigma^2}{2\pi \left| \sum_{j=0}^{p} \alpha_j (i\omega)^j \right|^2} \qquad -\pi < \omega < \pi \tag{2.3.1}$$

and

$$\gamma(h) = \int_{-\infty}^{\infty} e^{ih\omega} f(\omega) d\omega.$$

Now let $\mathcal{C}(z) = \sum_{j=0}^{p} \alpha_j z^j = 0$ with $\alpha_p = 1$ have q distinct roots $\lambda_1, \cdots, \lambda_q$ where λ_i has multiplicity m_i. Using contour integration, Doob (1953, p.543) showed that

$$\gamma(h) = \sigma^2 \sum_{i=1}^{q} c_i(h) e^{\lambda_i |h|} \tag{2.3.2}$$

where $c_i(h)$ is a polynomial in h of order m_i. Where all the roots are distinct ($m_i = 1, \forall i$), Jones (1981) gives

$$c_i(h) = \left[-2\mathrm{Re}(\lambda_i) \prod_{\substack{l=1 \\ l \neq i}}^{p} (\lambda_l - \lambda_i)(\bar{\lambda}_l + \lambda_i) \right]^{-1}$$

where $\bar{\lambda}_l$ denotes the complex conjugate of λ_l.

2.3.2 The derivative covariance function

The autocovariance function has proved useful in analysing discrete-time processes. Therefore, with continuous-time processes, it may be useful to define an analogous function replacing $X(t-j)$ by $X^{(j)}(t)$. We shall call this function the *derivative covariance function* or DCVF.

DEFINITION 2.1 *The derivative covariance function is defined as*

$$D_{j,k}(h) := \text{Cov}\left[X^{(j)}(t+h), X^{(k)}(t)\right], \qquad 0 \leq j,k \leq p-1 \tag{2.3.3}$$

Some properties of $D_{j,k}(h)$ follow immediately from its definition.

1 $D_{0,0}(h) = \gamma(h)$;

2 In general, $D_{j,k}(h) \neq D_{j,k}(-h)$;

3 $D_{j,k}(h) = D_{k,j}(-h)$;

The following theorem shows that the DCVF is closely related to the ACVF. The first part of this theorem is also proved by Ash and Gardner (1975, pp.263–264).

THEOREM 2.3.1 *If $\{X(t)\}$ is a CAR(p) process, then, for $h \geq 0$ and $0 \leq j,k \leq p-1$,*

$$D_{j,k}(h) = (-1)^k \gamma^{(j+k)}(h) \tag{2.3.4}$$

$$\textit{and} \quad D_{j,k}(-h) = (-1)^{j+k} D_{j,k}(h) = (-1)^j \gamma^{(j+k)}(h) \tag{2.3.5}$$

where $\gamma^{(j+k)}(h)$ denotes the $(j+k)$th derivative of $\gamma(\cdot)$ at h.

PROOF: I shall use proof by induction. Equation (2.3.4) is true for $j = k = 0$ by definition. Assume it is for $j = m < p-1$ and $k = n < p-1$. Then $\frac{d}{dh} D_{m,n}(h) = (-1)^n \gamma^{(m+n+1)}(h)$. Now,

$$\frac{d}{dh} D_{j,k}(h) = \frac{d}{dh} E\left[X^{(j)}(t+h) X^{(k)}(t)\right]$$

$$= E\left[X^{(j+1)}(t+h)X^{(k)}(t)\right] \quad \text{using Prabhu's (1965) Theorem 1.5, and the fact that } X(t) \text{ is } j+1 \text{ times differentiable}$$

$$= D_{j+1,k}(h).$$

Similarly,

$$\frac{d}{dh}D_{j,k}(h) = -D_{j,k+1}(h).$$

Hence,

$$\begin{aligned} D_{m+1,n}(h) &= (-1)^n\gamma^{(m+n+1)}(h) \\ D_{m,n+1}(h) &= (-1)^{n+1}\gamma^{(m+n+1)}(h). \end{aligned}$$

So by induction, (2.3.4) is true for all $j, k \leq p-1$. The second equation is obtained using the relation $D_{j,k}(h) = D_{k,j}(-h)$.

□

This result may be extended by defining $D_{j,p}(h)$ using the Itô integrals

$$D_{j,p}(h) := \lim_{T\to\infty}\frac{1}{T}\int_0^T X_*^{(j)}(t+h)\, d\left[X_*^{(p-1)}(t)\right] \tag{2.3.6}$$

where the limit exists in the sense of mean square convergence. Then $D_{j,p}(h) = -D_{j+1,p-1}(h) = (-1)^p\gamma^{(j+p)}(h)$, $j \leq p-2$. Similarly, for $h \neq 0$, $D_{p-1,p}(h) = (-1)^p\gamma^{(2p-1)}(h)$.

Difficulties arise with $D_{p-1,p}(0)$ since $\gamma^{(2p-2)}(h)$ is not differentiable at $h = 0$ (Doob, 1953, p.544). For example, in the case $p = 1$, the autocovariance function is $\gamma(h) = \frac{\sigma^2}{2\alpha_0}e^{-\alpha_0|h|}$ and so the first derivative is not defined in the usual sense at $h = 0$. This reflects the fact that in this case, $\{X(t)\}$ is not mean-square differentiable.

Nevertheless, Itô's formula provides a method of calculating $D_{p-1,p}(0)$. Define $Y[\boldsymbol{S}_*(t),t] = \frac{1}{2}\left[\boldsymbol{b}^T\boldsymbol{S}_*(t)\right]^2$ where $\boldsymbol{b} = [0,\cdots,0,1]^T$. Then $Y[\boldsymbol{S}_*(t),t] = \frac{1}{2}\left[X_*^{(p-1)}(t)\right]^2$. Using the multi-dimensional Itô formula (Øksendal, 1989, p.33),

$$dY[\boldsymbol{S}_*(t),t] = X_*^{(p-1)}(t)\,d\left[X_*^{(p-1)}(t)\right] + \frac{\sigma^2}{2}dt.$$

So

$$\begin{aligned}\frac{1}{T}\int_0^T X_*^{(p-1)}(t)\,d\left[X_*^{(p-1)}(t)\right] &= \frac{1}{T}\left\{Y[\boldsymbol{S}_*(T),T] - Y[\boldsymbol{S}_*(0),0]\right\} - \frac{\sigma^2}{2}\\ &= \frac{1}{2T}\left\{X_*^{(p-1)}(T)^2 - X_*^{(p-1)}(0)^2\right\} - \frac{\sigma^2}{2}.\end{aligned}$$

Thus, using (2.3.6),

$$D_{p-1,p}(0) = \frac{-\sigma^2}{2} \tag{2.3.7}$$

which is the right derivative of $-D_{p-1,p-1}(h)$ at $h = 0$ (Doob, 1953, p.544).

LEMMA 2.3.2 *If $\{X(t)\}$ is a CAR(p) process then, for $0 \le j,k \le p-1$,*

$$\begin{aligned} D_{j,k}(0) &= 0 \qquad && \text{if } j+k \text{ is odd;}\\ sign(D_{j,k}(0)) &= (-1)^{(3k+j)/2} && \text{if } j+k \text{ is even.}\end{aligned}$$

PROOF: Now $D_{j,k}(0) = D_{k,j}(0)$ by definition. Therefore, from Theorem 2.3.1, $(-1)^k\gamma^{(j+k)}(0) = (-1)^j\gamma^{(j+k)}(0)$. So, if $j+k$ is odd, then $(-1)^j \ne (-1)^k$ and $\gamma^{(j+k)}(0) = 0$. Hence, using Theorem 2.3.1 again, we obtain the first equation.

For even n where $n \le 2p-2$, $\gamma^{(n)}(0) = (-1)^{n/2}\text{Var}\left(X^{(n/2)}(t)\right)$, and so $\text{sign}(\gamma^{(n)}(0)) = (-1)^{n/2}$. In particular, $\text{sign}(\gamma^{(j+k)}(0)) = (-1)^{(j+k)/2}$ if $j+k$

is even. Hence, from Theorem 2.3.1, $\text{sign}(D_{j,k}(0)) = (-1)^k(-1)^{(j+k)/2} = (-1)^{(3k+j)/2}$.

□

2.3.3 Covariance matrix

Let $\Gamma_p(h) = \text{Cov}[\boldsymbol{S}(t), \boldsymbol{S}(t-h)]$ denote the autocovariance function of the stationary p-variate process $\{\boldsymbol{S}(t)\}$ which has jth component $X^{(j-1)}(t)$, $j = 1, \cdots, p$. Then the (j,k)th element of $\Gamma_p(h)$ is $D_{j-1,k-1}(h)$. Hence, for $h \geq 0$,

$$\Gamma_p(h) = \begin{pmatrix} \gamma^{(0)}(h) & -\gamma^{(1)}(h) & \cdots & (-1)^{p-1}\gamma^{(p-1)}(h) \\ \gamma^{(1)}(h) & -\gamma^{(2)}(h) & \cdots & (-1)^{p-1}\gamma^{(p)}(h) \\ \vdots & \vdots & & \vdots \\ \gamma^{(p-1)}(h) & -\gamma^{(p)}(h) & \cdots & (-1)^{p-1}\gamma^{(2p-2)}(h) \end{pmatrix}$$

and $\Gamma_p(-h) = \Gamma_p^T(h)$ using Theorem 2.3.1. Note that in the case $h = 0$ every second element of this matrix is zero according to Lemma 2.3.2.

Multiplying both sides of (2.1.3) by $\boldsymbol{S}_*^T(t-h)$ and taking expectations, we obtain

$$d\Gamma_p(h) = A\Gamma_p(h)dh. \tag{2.3.8}$$

Thus, using (2.1.4),

$$\Gamma_p(h) = e^{Ah}\Gamma_p(0) = \sigma^2 \int_0^\infty e^{A(u+h)}\boldsymbol{b}\boldsymbol{b}^T e^{A^T u}du. \tag{2.3.9}$$

THEOREM 2.3.3 *If $\{X(t)\}$ is a CAR(p) process, then $\Gamma_p(0)$ is strictly positive definite.*

PROOF: $\Gamma_p(0)$ is non-negative definite since it is a covariance matrix. It remains to show that $\boldsymbol{c}^T\Gamma_p(0)\boldsymbol{c} \neq 0$ for any $\boldsymbol{c} \neq \mathbf{0}$.

Suppose there exists $\boldsymbol{c} \neq \mathbf{0}$ such that $\boldsymbol{c}^T\Gamma_p(0)\boldsymbol{c} = 0$. Then, from (2.3.9),

$$\boldsymbol{c}^T\Gamma_p(0)\boldsymbol{c} = \sigma^2 \int_0^\infty \boldsymbol{c}^T e^{-Au}\boldsymbol{b}\boldsymbol{b}^T e^{-A^T u}\boldsymbol{c}\, du = 0$$

which means that $\boldsymbol{c}^T e^{-Au}\boldsymbol{b} = 0$ for all $u \geq 0$. But, there exists a $u \geq 0$ such that $e^{-Au}\boldsymbol{b} = \boldsymbol{c}$, and for this u, $\boldsymbol{c}^T e^{-Au}\boldsymbol{b} = \boldsymbol{c}^T\boldsymbol{c} > 0$ since $\boldsymbol{c} \neq \mathbf{0}$. Hence, the assumption leads to a contradiction.

□

A corollary of this theorem is that $|\Gamma_p(0)| > 0$. Evaluating $|\Gamma_p(0)|$ leads to restrictions on $\gamma^{(k)}(0)$. For $p = 1$ and $p = 2$, these restrictions are trivial and may be obtained from Lemma 2.3.2. The restrictions for $p = 3$ and $p = 4$ are:

$$\begin{aligned} p = 3 \quad & \left(\gamma^{(0)}(0)\gamma^{(4)}(0) - \left[\gamma^{(2)}(0)\right]^2\right) > 0; \text{ and} \\ p = 4 \quad & \left(\gamma^{(0)}(0)\gamma^{(4)}(0) - \left[\gamma^{(2)}(0)\right]^2\right)\left(\gamma^{(2)}(0)\gamma^{(6)}(0) - \left[\gamma^{(4)}(0)\right]^2\right) > 0. \end{aligned}$$

2.4 Yule–Walker Equations for CAR Processes

Let $\{X(t)\}$ be the CAR(p) process defined by (2.1.3) and (2.1.4). Then, multiplying equation (2.1.1) by $X_*^{(j)}(t+h)$ and taking expectations gives

$$\alpha_0 D_{j,0}(h) + \alpha_1 D_{j,1}(h) + \cdots + \alpha_p D_{j,p}(h) = 0 \qquad j = 0, 1, \ldots, p-1. \quad (2.4.1)$$

These are the Yule–Walker equations for CAR processes. They are analogous to the traditional discrete time Yule–Walker equations which can be obtained in a similar way replacing $X^{(j)}(t)$ by $X(t-j)$ and letting $h = 0$.

Let $h \geq 0$. Then substituting $(-1)^k \gamma^{(j+k)}(h)$ for $D_{j,k}(h)$ into (2.4.1) gives

$$(-1)^p \gamma^{(j+p)}(h) + (-1)^{p-1} \alpha_{p-1} \gamma^{(j+p-1)}(h) + \cdots + \alpha_0 \gamma^{(j)}(h) = 0 \qquad h \geq 0 \tag{2.4.2}$$

where $j = 0, 1, \cdots, p-1$ and $\gamma^{(2p-1)}(0)$ is to be interpreted as the right derivative of $\gamma^{(2p-2)}(h)$ at $h = 0$.

Similarly, for $h \leq 0$, substitute $(-1)^j \gamma^{(j+k)}(-h)$ for $D_{j,k}(h)$ into (2.4.1) to obtain

$$\gamma^{(j+p)}(h) + \alpha_{p-1} \gamma^{(j+p-1)}(h) + \cdots + \alpha_0 \gamma^{(j)}(h) = 0 \qquad h \geq 0. \tag{2.4.3}$$

Note that these equations are also obtained from the last row of (2.3.8).

It will be convenient to write (2.4.2) in the matrix form

$$\Gamma_p(h)\boldsymbol{\alpha} + \boldsymbol{\gamma}_p(h) = 0 \tag{2.4.4}$$

where

$$\boldsymbol{\gamma}_p(h) = (-1)^p \left[\gamma^{(p)}(h), \gamma^{(p+1)}(h), \cdots, \gamma^{(2p-1)}(h)\right]^T$$

$$\text{and} \qquad \boldsymbol{\alpha} = [\alpha_0, \alpha_1, \cdots, \alpha_{p-1}]^T.$$

Note that $\Gamma_p(0)$ is non-singular by Theorem 2.3.3. So, for $h = 0$ we may write

$$\boldsymbol{\alpha} = -\Gamma_p^{-1}(0)\boldsymbol{\gamma}_p(0). \tag{2.4.5}$$

2.4.1 More stationarity conditions

If the process is stationary, we may combine the Yule–Walker equations with Corollary 2.2.4 to obtain conditions on $\gamma^{(k)}(0)$.

COROLLARY 2.4.1 *Let $\boldsymbol{S}(t)$ satisfy (2.1.3) and (2.1.4) and let $X(t)$ be the first component of $\boldsymbol{S}(t)$. If $X(t)$ is stationary then the following conditions hold.*

$$\begin{aligned}
&p = 1 \quad \gamma^{(1)}(0) < 0; \\
&p = 2 \quad \gamma^{(3)}(0) > 0; \\
&p = 3 \quad \gamma^{(5)}(0) < 0 \quad \textit{and} \quad \left(\gamma^{(0)}(0)\gamma^{(4)}(0) - \left[\gamma^{(2)}(0)\right]^2\right) > 0; \textit{ and} \\
&p = 4 \quad \gamma^{(7)}(0) > 0, \left(\gamma^{(0)}(0)\gamma^{(4)}(0) - \left[\gamma^{(2)}(0)\right]^2\right) > 0 \\
&\qquad\qquad \textit{and} \quad \left(\gamma^{(2)}(0)\gamma^{(6)}(0) - \left[\gamma^{(4)}(0)\right]^2\right) > 0.
\end{aligned}$$

Here $\gamma^{(2p-1)}(0)$ denotes the right derivative of $\gamma^{(2p-2)}(h)$ at $h = 0$ or $(-1)^p D_{p-1,p}(0)$. Thus, using (2.3.7), the first inequality given in each of these cases may be restated as $\sigma^2 > 0$. Compare these with the restrictions given at the end of the previous section.

2.5 Regularly Sampled CAR Processes

Let $\{X(t)\}$ denote a CAR(1) process. Then, from (2.3.2), the autocovariance function of $\{X(t)\}$ is

$$\gamma(s) = \frac{\sigma^2}{2\alpha_0} e^{-\alpha_0|s|}.$$

Thus, if $\{X_t\}$ denotes observations from a CAR(1) process which is sampled at intervals of Δ, the autocovariance function of $\{X_t\}$ is given by

$$\gamma_k = \gamma(k\Delta) = \frac{\sigma^2}{2\alpha_0} e^{-k\alpha_0\Delta}, \qquad k = 0, 1, 2, \ldots. \tag{2.5.1}$$

Therefore, $\{X_t\}$ is the discrete time AR(1) process,

$$X_t = -\frac{\beta}{\alpha_0}\left(1 - e^{-\alpha_0\Delta}\right) + e^{-\alpha_0\Delta} X_{t-1} + e_t \tag{2.5.2}$$

where e_t is white noise with variance

$$\sigma_e^2 = \gamma_0\left(1 - e^{-2\alpha_0\Delta}\right) = \frac{\sigma^2}{2\alpha_0}\left(1 - e^{-2\alpha_0\Delta}\right).$$

Conversely, any discrete time AR(1) with positive coefficient can be embedded in a CAR(1). The same is not true for the discrete time AR(1) process, $X_t = \phi X_{t-1} + e_t$, where $\phi \leq 0$. For a discussion of this problem and the more general problem of embedding a discrete time ARMA(p, q) with $q < p$ in a continuous time ARMA(p', q') with $q' < p'$, see Chan and Tong (1987) and He and Wang (1989).

With a little more difficulty, Pandit and Wu (1983, pp.260–267) derive the autocovariance function of a sample taken at constant time steps from a CAR(2) process and thereby show that the sampled process is an ARMA(2,1). Again, some restrictions are required on the ARMA(2,1) parameters and so the general ARMA(2,1) process can not always be embedded in a CAR(2) process.

Generalising these results leads to the following theorem.

THEOREM 2.5.1 *If a CAR(p) process is sampled at regular time intervals,*

the resulting observations follow a discrete time ARMA$(p, p-1)$ process.

Bartlett (1946) proved this result for the case $p = 2$. The more general case was proved by Phillips (1959) using an argument based on the spectra of the discrete and continuous processes. Phadke and Wu (1974) follow a similar approach to Bartlett (1946) and consider the general CAR(p) model. They prove the result by showing that when a discrete time AR(p) process with specified coefficients is fitted to a sample from a continuous time CAR(p) process, the residuals have non-zero autocorrelations up to lag $p-1$.

The actual parameters in the resulting ARMA$(p, p-1)$ model are, in general, complicated functions of the parameters of the CAR(p) model. Details are given in Phadke and Wu (1974) and Pandit and Wu (1983) for the cases $p = 1$ and $p = 2$. Phillips (1959) derives the general result.

Chapter 3

CONTINUOUS TIME LINEAR AUTOREGRESSIVE MODELLING

In this chapter, I consider the problem of estimating a CAR model from observations.

The problem of estimating the model parameters has been the subject of much statistical, econometric and engineering literature. For the estimation of the model from a continuous record, see Bartlett (1946, 1955), Dzhaparidze (1971) and Priestley (1981). In practice, a sample function is observed discretely at the points $t_0 < t_1 < \cdots < t_N$. For previous work in the case of equally spaced data ($t_{i+1} - t_i = \Delta$ for all i), see, for example, Bartlett (1946), Durbin (1961), Phillips (1972) and Bergstrom (1984). These exploit the well-known result (discussed in Section 2.5) that an equally spaced sample from a CAR(p) process is itself an ARMA(p,$p-1$) process. The case with unequally spaced observations where the times of observation are non-random has been considered by Jones (1981), Robinson (1977) and others. Robinson (1980) considers the case where the times of observation form a stationary point process.

In Section 3.1, I derive a new approach to estimating CAR models, based on the Yule–Walker equations of the previous chapter. For continuously recorded data, it is shown that these estimators are least squares estimators and have the same asymptotic distribution as maximum likelihood estimators. Conditions under which the estimated models are stationary are also derived. In practice, though, data can only be observed discretely. For discrete data, I consider approximations to the continuous time estimators. It is shown that some of these discrete time estimators are asymptotically biased. Alternative estimators based on the autocovariance function are suggested. These are asymptotically unbiased and are a fast alternative to the maximum likelihood estimators described by Jones (1981). They may also be used as starting values for maximum likelihood estimation.

Section 3.2 describes the maximum likelihood approach to estimating CAR models from discrete observations at deterministic times. The presentation given here is based largely on Jones (1981).

A modelling paradigm for CAR processes is outlined in Section 3.3, including order selection, model estimation and diagnostic checks. In Section 3.4, this is applied to three real time series: the IBM closing stock price data, Wolf's sunspot numbers and the Canadian lynx data.

3.1 Yule–Walker Estimators

In this section, I consider a continuous time analogue of the Yule–Walker method for estimating discrete time autoregressive models. Following the Yule–Walker approach, I shall use the Yule–Walker equations developed in the

previous chapter to estimate the model coefficients by replacing the covariances by sample estimates based on continuously recorded data. The Yule–Walker estimators are shown to be least squares estimators which converge in distribution to maximum likelihood estimators, thus maintaining the close analogy between the continuous time Yule–Walker estimators and the discrete time Yule–Walker estimators.

It is assumed throughout this section that data have been mean-corrected where necessary. Thus, we fit models with $\beta = 0$.

3.1.1 Estimating the DCF

Let $\{x(t)\}$ denote a sample function of a CAR(p) process, $\{X(t)\}$, and suppose we are able to observe and record it continuously without error between the times 0 and T. I now consider the problem of estimating $D_{j,k}(0)$, $0 \leq j \leq p-1$, $0 \leq k \leq p$, based on $\{x(t)\}$.

For $j = 0, 1, \ldots, p-1$, $X^{(j)}(t)$ is equivalent to a process with almost all sample functions continuous (Jazwinski, 1970, p.111). Therefore, $\{X(t)\}$ is equivalent to a process with almost all sample functions possessing continuous derivatives up to order $p-1$. Hence, $x^{(j)}(t)$ $(0 \leq j \leq p-1)$ may be calculated from the data.

Then, $D_{j,k}(0)$, $0 \leq j \leq p-1$, $0 \leq k \leq p$ may be estimated as the sample covariances

$$\hat{D}_{j,k} = \frac{1}{T}\int_0^T x^{(j)}(t)\, x^{(k)}(t)dt, \qquad k < p, \tag{3.1.1}$$

$$\text{and} \quad \hat{D}_{j,p} = \frac{1}{T}\int_0^T x^{(j)}(t)\, dx^{(p-1)}(t) \tag{3.1.2}$$

Then $\hat{D}_{j,k} \xrightarrow{p} D_{j,k}(0)$ as $T \to \infty$ (Yaglom, 1987, pp.231–233).

3.1.2 Yule–Walker estimators

Consider equation (2.4.5) and let $\Gamma_{j,k}$ denote the (j,k)th element of $\Gamma_p(0)$ and γ_j denote the jth element of $\boldsymbol{\gamma}_p(0)$ $(j,k = 1,\ldots,p)$. Then, replacing $\Gamma_{j,k}$ by the sample estimate $\hat{D}_{j-1,k-1}$ and γ_j by the sample estimate $\hat{D}_{j-1,p}$, we obtain the Yule–Walker estimator

$$\hat{\boldsymbol{\alpha}} = -\hat{\Gamma}_p^{-1}(0)\hat{\boldsymbol{\gamma}}_p(0). \tag{3.1.3}$$

Bartlett (1946, 1955) and Priestley (1981) derive these estimators in the cases $p = 1$ and $p = 2$ using a least squares approach. The following theorem shows that the Yule–Walker estimates satisfy least squares criteria for all p.

THEOREM 3.1.1 *Let $\{x(t)\}$ be a sample function from a CAR(p) process observed at time $0 \le t \le T$. Then the least squares estimators of $\boldsymbol{\alpha}$, obtained by minimising*

$$Q(\boldsymbol{\alpha}) = \int_0^T [\sigma dW(t)]^2 = \int_0^T \left[dx^{(p-1)}(t) + \alpha_{p-1}x^{(p-1)}(t)dt + \cdots + \alpha_0 x(t)dt\right]^2 \tag{3.1.4}$$

are the Yule–Walker estimators given by (3.1.3).

PROOF: Differentiating (3.1.4) with respect to α_j gives

$$\int_0^T x^{(j)}(t)\left[dx^{(p-1)}(t) + \breve{\alpha}_{p-1}x^{(p-1)}(t)dt + \cdots + \breve{\alpha}_0 x(t)dt\right] = 0,$$

where $j = 0, \ldots, p-1$ and $\check{\alpha}_j$ is the least squares estimator of α_i. Equivalently,

$$\check{\alpha}_0 \hat{D}_{j,0} + \check{\alpha}_1 \hat{D}_{j,1} + \cdots + \check{\alpha}_{p-1} \hat{D}_{j,p-1} = -\hat{D}_{j,p} \qquad j = 0, 1, \ldots, p-1,$$

which are the Yule–Walker equations with $D_{j,k}$ replaced by $\hat{D}_{j,k}$. Hence $\check{\alpha}_j \equiv \hat{\alpha}_j$ for all j.

□

The asymptotic distribution of the Yule–Walker estimators is given in the following theorem.

THEOREM 3.1.2 *Let $\hat{\boldsymbol{\alpha}}$ denote the estimate of $\boldsymbol{\alpha}$ defined by (3.1.3). Then*

$$\sqrt{T}\,(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{d} N\left(\mathbf{0}\ ,\ \sigma^2 \Gamma_p^{-1}(0)\right) \qquad \text{as } T \to \infty.$$

PROOF: Let $\{x(t)\}$, $0 < t < T$ denote a sample function of a CAR(p) process. Also let n be some positive integer and define $\Delta = T/n$ and $t_i = i\Delta$. Then, for large n, $\hat{D}_{j,k}$ may be approximated by

$$\begin{aligned} \hat{D}_{j,k,n} &= \frac{1}{n} \sum_{i=0}^{n} x^{(j)}(t_i)\, x^{(k)}(t_i), \qquad k < p, \\ \text{and} \quad \hat{D}_{j,p,n} &= \frac{1}{n\Delta} \sum_{i=0}^{n-1} x^{(j)}(t_i) \left[x^{(p-1)}(t_{i+1}) - x^{(p-1)}(t_i)\right] \end{aligned}$$

so that $\hat{D}_{j,k,n} \xrightarrow{p} \hat{D}_{j,k}$, $0 \le k \le p$ as $n \to \infty$ (Yaglom, 1987).

Define $\hat{\boldsymbol{\alpha}}_n$ to be the estimator obtained by replacing $\hat{D}_{j,k}$ by $\hat{D}_{j,k,n}$ in (3.1.3). Then $\hat{\boldsymbol{\alpha}}_n$ is the standard least squares estimator for the regression problem

$$Y(t_i) = -\sqrt{\Delta}\left(\alpha_0 x(t_i) + \cdots + \alpha_{p-1} x^{(p-1)}(t_i)\right) + e(t_i)$$

$$\text{where} \quad Y(t_i) = \left[x^{(p-1)}(t_{i+1}) - x^{(p-1)}(t_i)\right]/\sqrt{\Delta}$$
$$\text{and} \quad e(t_i) = \sigma\left[W(t_{i+1}) - W(t_i)\right]/\sqrt{\Delta}.$$

That is

$$\boldsymbol{Y} = -X\boldsymbol{\alpha} + \boldsymbol{e} \qquad \text{and} \qquad \hat{\boldsymbol{\alpha}}_n = -(X^TX)^{-1}(X^T\boldsymbol{Y})$$

$$\text{where} \quad \boldsymbol{Y} = [Y(t_0), \ldots, Y(t_{n-1})]^T, \qquad \boldsymbol{e} = [e(t_0), \ldots, e(t_{n-1})]^T$$

$$\text{and} \quad X = \sqrt{\Delta}\begin{bmatrix} x(t_0) & x^{(1)}(t_0) & \cdots & x^{(p-1)}(t_0) \\ x(t_1) & x^{(1)}(t_1) & \cdots & x^{(p-1)}(t_1) \\ \vdots & \vdots & & \vdots \\ x(t_{n-1}) & x^{(1)}(t_{n-1}) & \cdots & x^{(p-1)}(t_{n-1}) \end{bmatrix}.$$

Thus, $e(t_i)$ is a sequence of independent, normally distributed random variables with zero mean and variance σ^2.

Consider

$$T^{\frac{1}{2}}(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}) = T^{\frac{1}{2}}\left[(X^TX)^{-1}X^T(X\boldsymbol{\alpha} - \boldsymbol{e}) - \boldsymbol{\alpha}\right] = -T(X^TX)^{-1}\left[T^{-\frac{1}{2}}X^T\boldsymbol{e}\right].$$

By setting $\boldsymbol{U}(t_i) = \left[x(t_i), x^{(1)}(t_i), \ldots, x^{(p-1)}(t_i)\right]^T e(t_i)$ we have

$$T^{-\frac{1}{2}}X^T\boldsymbol{e} = T^{-\frac{1}{2}}\sum_{i=0}^{n-1}\sqrt{\Delta}\boldsymbol{U}(t_i) \xrightarrow{p} n^{-\frac{1}{2}}\sum_{i=0}^{n}\boldsymbol{U}(t_i) \qquad \text{as } n \to \infty.$$

Also, $E\left[\boldsymbol{U}(t_i)\right] = \boldsymbol{0}$, $\text{Cov}\left[\boldsymbol{U}(t_i), \boldsymbol{U}(t_j)\right] = O_{p\times p}$, $i \neq j$ and $\text{Cov}\left[\boldsymbol{U}(t_i), \boldsymbol{U}(t_i)\right] = \sigma^2\Gamma_p(0)$ since $W(t_i)$ and $W(t_{i+1})$ are independent of $x(t_i), \ldots, x^{(p-1)}(t_i)$.

Now $x^{(k)}(t)$ may be written as $\sigma\int_0^\infty \psi_k(y)\,dW(t-y)$ (see Arató, 1982, p.119). Hence, we may define the sequence of approximations $x_m^{(k)}(t_i) = \sigma\int_0^{t_m}\psi_k(y)\,dW(t_i - y)$ where m is a fixed positive integer. Let $\boldsymbol{S}_m(t_i) = \left[x_m(t_i), x_m^{(1)}(t_i), \ldots, x_m^{(p-1)}(t_i)\right]^T$ and $\boldsymbol{U}_m(t_i) = \boldsymbol{S}_m(t_i)e(t_i)$ and let $\boldsymbol{\lambda}$ be a fixed

but arbitrary vector in $\mathbb{R}^p$. Then $Z_i = \boldsymbol{\lambda}^T \boldsymbol{U}_m(t_i)$ is a strictly stationary sequence with zero mean and variance $\sigma^2 \boldsymbol{\lambda}^T \Gamma_{m,p}(0) \boldsymbol{\lambda}$ where $\Gamma_{m,p}(0)$ is the covariance matrix of $\boldsymbol{S}_m(t)$. The sequence Z_i also has the property $E[Z_i Z_{i+h}] = 0$ when $h > m+1$. That is, it is $(m+1)$-dependent. Hence, we may apply the central limit theorem for dependent stationary sequences (see, e.g., Brockwell and Davis, 1991, p.213) to obtain

$$n^{-\frac{1}{2}} \sum_{i=0}^{n-1} \boldsymbol{\lambda}^T \boldsymbol{U}_m(t_i) \xrightarrow{d} \boldsymbol{\lambda}^T \boldsymbol{V}_m \quad \text{where} \quad \boldsymbol{V}_m \stackrel{d}{=} N\left(\mathbf{0}\ ,\ \sigma^2 \Gamma_{m,p}(0)\right).$$

Since $\sigma^2 \Gamma_{m,p}(0) \to \sigma^2 \Gamma_p(0)$ as $m \to \infty$, we have $\boldsymbol{\lambda}^T \boldsymbol{V}_m \xrightarrow{d} \boldsymbol{\lambda}^T \boldsymbol{V}$ where $\boldsymbol{V} \stackrel{d}{=} N(\mathbf{0}\ ,\ \sigma^2 \Gamma_p(0))$. Also, $n^{-1}\mathrm{Var}\left(\boldsymbol{\lambda}^T \sum_{i=0}^{n-1} [\boldsymbol{U}_m(t_i) - \boldsymbol{U}(t_i)]\right) \to 0$ as $m \to \infty$.

Since $x_m(t) \xrightarrow{ms} x(t)$, application of Brockwell and Davis's Proposition 6.3.9 (1991, p.207–208) and the Cramer–Wold device gives

$$T^{-\frac{1}{2}} X^T \boldsymbol{e} \xrightarrow{d} N\left(\mathbf{0}\ ,\ \sigma^2 \Gamma_p(0)\right).$$

It follows from (3.1.1) that $T(X^T X)^{-1} \xrightarrow{p} \Gamma_p^{-1}(0)$. Then, using Brockwell and Davis's Proposition 6.4.2 (1991, p.211),

$$\hat{\boldsymbol{\alpha}}_n \stackrel{d}{=} AN\left(\boldsymbol{\alpha}\ ,\ T^{-1}\sigma^2 \Gamma_p^{-1}(0)\right).$$

Now $\hat{\boldsymbol{\alpha}}_n \xrightarrow{p} \hat{\boldsymbol{\alpha}}$ as $n \to \infty$ since $D_{j,k,n} \xrightarrow{p} D_{j,k}$. So

$$\hat{\boldsymbol{\alpha}} \stackrel{d}{=} AN\left(\boldsymbol{\alpha}\ ,\ T^{-1}\sigma^2 \Gamma_p^{-1}(0)\right).$$

□

Arató (1982, p.261) shows that the maximum likelihood estimator of $\boldsymbol{\alpha}$ for a continuously recorded process between 0 and T is asymptotically normally

distributed with mean $\boldsymbol{\alpha}$ and covariance matrix $\frac{1}{T}\sigma^2\Gamma_p^{-1}(0)$. Thus, the Yule–Walker estimator has the same asymptotic distribution as the maximum likelihood estimator.

The results of Section 2.3.2 suggest some improvements to the Yule–Walker estimates. I shall use a bar ($\bar{\ }$) to distinguish the improved estimates from those above denoted by a hat ($\hat{\ }$).

Using Theorem 2.3.1, we define

$$\begin{aligned} \bar{\gamma}^{(2n)}(0) &:= (-1)^n \hat{D}_{n,n} \\ \bar{\Gamma}_{j,k} &:= (-1)^{k-1}\bar{\gamma}^{(j+k-2)}(0) \qquad \text{for } j+k \text{ even} \\ \text{and} \qquad \bar{\Gamma}_{j,k} &:= 0 \qquad\qquad\qquad\qquad\quad \text{for } j+k \text{ odd.} \end{aligned}$$

The last definition is based on Lemma 2.3.2. Similarly, define

$$\begin{aligned} \bar{\gamma}_j &:= (-1)^p\bar{\gamma}^{(j+p-1)}(0) \qquad\qquad \text{for } j+p-1 \text{ even} \\ \bar{\gamma}_j &:= 0 \qquad\qquad\qquad\qquad\quad \text{for } j+p-1 < 2p \text{ odd} \\ \text{and} \qquad \bar{\gamma}_p &:= (-1)^p\bar{\gamma}^{(2p-1)}(0) := \hat{D}_{p-1,p} \end{aligned}$$

Then (3.1.3) may be replaced by

$$\bar{\boldsymbol{\alpha}} = -\bar{\Gamma}_p^{-1}(0)\bar{\boldsymbol{\gamma}}_p(0). \tag{3.1.5}$$

Consider the structure of $\bar{\Gamma}_p(0)$. Using the above definitions,

$$\bar{\Gamma}_p(0) = \begin{pmatrix} \bar{\gamma}^{(0)}(0) & 0 & \bar{\gamma}^{(2)}(0) & 0 & \bar{\gamma}^{(4)}(0) & \cdots \\ 0 & -\bar{\gamma}^{(2)}(0) & 0 & -\bar{\gamma}^{(4)}(0) & \cdots & \\ \bar{\gamma}^{(2)}(0) & 0 & \bar{\gamma}^{(4)}(0) & \ddots & & \\ 0 & -\bar{\gamma}^{(4)}(0) & \ddots & & & \\ \bar{\gamma}^{(4)}(0) & \vdots & & & & \\ \vdots & & & & & \end{pmatrix}. \tag{3.1.6}$$

Separating the odd and even rows of $\bar{\Gamma}_p(0)$, we can write equation (3.1.5) as

$$\bar{\boldsymbol{\alpha}}_0 = -F_p^{-1}\boldsymbol{f}_p \tag{3.1.7}$$

$$\text{and} \qquad \bar{\boldsymbol{\alpha}}_1 = -G_p^{-1}\boldsymbol{g}_p \tag{3.1.8}$$

where

$$F_p = \begin{pmatrix} \bar{\gamma}^{(0)}(0) & \bar{\gamma}^{(2)}(0) & \cdots & \bar{\gamma}^{(f)}(0) \\ \bar{\gamma}^{(2)}(0) & \bar{\gamma}^{(4)}(0) & \cdots & \bar{\gamma}^{(f+2)}(0) \\ \vdots & \vdots & & \vdots \\ \bar{\gamma}^{(f)}(0) & \bar{\gamma}^{(f+2)}(0) & \cdots & \bar{\gamma}^{(2f)}(0) \end{pmatrix}, \qquad \boldsymbol{f}_p = \begin{pmatrix} \bar{\gamma}^{(p)}(0) \\ \bar{\gamma}^{(p+2)}(0) \\ \vdots \\ \bar{\gamma}^{(p+f)}(0) \end{pmatrix},$$

$$G_p = \begin{pmatrix} \bar{\gamma}^{(2)}(0) & \bar{\gamma}^{(4)}(0) & \cdots & \bar{\gamma}^{(g)}(0) \\ \bar{\gamma}^{(4)}(0) & \bar{\gamma}^{(6)}(0) & \cdots & \bar{\gamma}^{(g+2)}(0) \\ \vdots & \vdots & & \vdots \\ \bar{\gamma}^{(g)}(0) & \bar{\gamma}^{(g+2)}(0) & \cdots & \bar{\gamma}^{(2g-2)}(0) \end{pmatrix}, \qquad \boldsymbol{g}_p = \begin{pmatrix} \bar{\gamma}^{(p+1)}(0) \\ \bar{\gamma}^{(p+3)}(0) \\ \vdots \\ \bar{\gamma}^{(p+g-1)}(0) \end{pmatrix},$$

$\bar{\boldsymbol{\alpha}}_0 = [\bar{\alpha}_0, \bar{\alpha}_2, \cdots, \bar{\alpha}_f]^T, \quad \bar{\boldsymbol{\alpha}}_1 = [\bar{\alpha}_1, \bar{\alpha}_3, \cdots, \bar{\alpha}_{g-1}]^T,$

$$f = \begin{cases} p-2 & p \text{ even;} \\ p-1 & p \text{ odd;} \end{cases} \qquad \text{and} \quad g = \begin{cases} p & p \text{ even;} \\ p-1 & p \text{ odd.} \end{cases}$$

Thus,

$$\begin{aligned} \text{for } p \text{ odd} \qquad & \boldsymbol{f}_p = [0, \cdots, 0, \bar{\gamma}^{(2p-1)}(0)]^T \\ \text{and for } p \text{ even} \qquad & \boldsymbol{g}_p = [0, \cdots, 0, \bar{\gamma}^{(2p-1)}(0)]^T. \end{aligned}$$

So instead of inverting a matrix of order $p \times p$ as in equation (3.1.3), one need only invert two matrices about half that size.

Clearly, $\bar{\boldsymbol{\alpha}} \xrightarrow{p} \hat{\boldsymbol{\alpha}}$ and so these have the same asymptotic distribution as $\hat{\boldsymbol{\alpha}}$.

One useful property of the improved Yule–Walker estimates is that they allow more rapid computation when estimating models of successively higher orders. In particular, note that if p is even, $F_p = F_{p-1}$ and if p is odd, $G_p = G_{p-1}$.

3.1.3 Stationarity conditions

Corollary 2.2.4 may be used to give necessary and sufficient conditions under which the estimated model is stationary.

THEOREM 3.1.3 *If the observed data are not all equal, the following conditions are necessary and sufficient for the estimated model to be stationary:*

$p = 1 \quad \hat{\gamma}^{(1)}(0) < 0;$

$p = 2 \quad \hat{\gamma}^{(3)}(0) > 0;$

$p = 3 \quad \hat{\gamma}^{(5)}(0) < 0 \quad \textit{and} \quad \left(\hat{\gamma}^{(0)}(0)\hat{\gamma}^{(4)}(0) - \left[\hat{\gamma}^{(2)}(0)\right]^2\right) > 0;$

$p = 4 \quad \hat{\gamma}^{(7)}(0) > 0, \quad \left(\hat{\gamma}^{(0)}(0)\hat{\gamma}^{(4)}(0) - \left[\hat{\gamma}^{(2)}(0)\right]^2\right) > 0$

$\textit{and} \quad \left(\hat{\gamma}^{(2)}(0)\hat{\gamma}^{(6)}(0) - \left[\hat{\gamma}^{(4)}(0)\right]^2\right) > 0.$

PROOF: For $k = 0, 1, \ldots, p$, the estimated derivatives, $\hat{\gamma}^{(2k)}(0)$, are calculated as the product of $(-1)^k$ and the sample variance of $\hat{x}^{(k)}(t)$. Now $\{\hat{x}^{(k)}(t_i)\}$ are not all equal if and only if the observations are not all equal. In this case, $\hat{\gamma}^{(2k)}(0)$ has sign $(-1)^k$.

So, in the case $p = 1$, $\hat{\alpha}_0 = -\hat{\gamma}^{(1)}(0)/\hat{\gamma}^{(0)}(0) > 0$ if and only if $\hat{\gamma}^{(1)}(0) < 0$. The result then follows from Corollary 2.2.4

The results for $p = 2$, 3 and 4 follow similarly by writing the conditions in Corollary 2.2.4 in terms of the estimates of $\gamma^{(k)}(0)$.

□

This 'necessary' part of this result may also have been obtained from Corollary 2.4.1 since, if the estimated model is stationary, the DCF of the model must satisfy the equations (2.4.1) and so coincides with the corresponding sample DCF at lag 0.

3.1.4 Estimating the instantaneous variance

In this section, I consider how to estimate the parameter σ^2.

For CAR models there is no direct analogue of the Yule–Walker estimate of residual variance. Nevertheless, it is possible to derive a quick and simple estimate based on the process variance $\gamma(0)$ and coefficients $\alpha_0, \ldots, \alpha_{p-1}$.

Equation (2.3.2) gives

$$\sigma^2 = \gamma(0) \left[\sum_{i=1}^{q} c_i(0) \right]^{-1} \tag{3.1.9}$$

where λ_i is a root of $\mathcal{C}(z) = \sum_{j=0}^{p} \alpha_p z^j = 0$ with multiplicity m_i and $c_i(0)$ is a

function of $\lambda_1, \ldots, \lambda_1$. If all the roots are distinct ($m_i = 1$, $\forall i$),

$$c_i(h) = \left[-2\mathrm{Re}(\lambda_i) \prod_{\substack{l=1 \\ l \neq i}}^{p} (\lambda_l - \lambda_i)(\bar{\lambda}_l + \lambda_i) \right]^{-1} \tag{3.1.10}$$

where $\bar{\lambda}_l$ denotes the complex conjugate of λ_l.

The roots λ_i may be estimated from the coefficient estimates and $\gamma(0)$ may be estimated from the data. Then (3.1.9) may be used to estimate σ^2.

For $p = 1$, this yields the simple estimate

$$\hat{\sigma}^2 = 2\hat{\alpha}_0 \hat{\gamma}(0)$$

and for $p = 2$,

$$\hat{\sigma}^2 = 2\hat{\alpha}_0 \hat{\alpha}_1 \hat{\gamma}(0).$$

In these cases, if the Yule–Walker estimates of the previous section are used, then $\hat{\sigma}^2 = -2\hat{D}_{p-1,p}$; a result which could have been derived from (2.3.7).

3.1.5 Yule–Walker estimates for discrete observations

In practice, observations are recorded at discrete times. So to use the above estimators, it is necessary to derive discrete time approximations to them.

Suppose the data consist of observations taken at times $0 = t_1 < t_2 < \cdots < t_N = T$ and let $\Delta_i = t_{i+1} - t_i$ and $\Delta = \sup_i \Delta_i$. Then, for small Δ, it seems reasonable to estimate $x^{(j)}(t_i)$ by numerical derivatives $\hat{x}^{(j)}(t_i)$ and $\hat{D}_{j,k}$ by the

numerical integrals

$$\tilde{D}_{j,k} = \frac{1}{T}\sum_{i=1}^{N} \hat{x}^{(j)}(t_i)\,\hat{x}^{(k)}(t_i)\Delta_i, \qquad k < p, \tag{3.1.11}$$

$$\text{and} \quad \tilde{D}_{j,p} = \frac{1}{T}\sum_{i=1}^{N} \hat{x}^{(j)}(t_i)\,\left[\hat{x}^{(p-1)}(t_{i+1}) - \hat{x}^{(p-1)}(t_i)\right] \tag{3.1.12}$$

Note that for $\tilde{D}_{j,p}$, it is crucial that $\hat{x}^{(j)}(t)$ is evaluated at the left hand end of the interval $[t_i, t_{i+1}]$. If it is evaluated at other values in the interval, the sum will converge to a different value. For example, if $\hat{x}^{(j)}(t)$ is evaluated at the mid-point of the interval $[t_i, t_{i+1}]$, $\tilde{D}_{j,p}$ will converge to the Stratonovich integral rather than the Itô integral (Øksendal, 1989, p.16).

Define a discrete form of the Yule–Walker estimators by replacing $\hat{D}_{j,k}$ by $\tilde{D}_{j,k}$ in (3.1.5). Let $\tilde{\boldsymbol{\alpha}}$ denote the estimators which are obtained in this way.

Now suppose the data are equally spaced ($\Delta_i = \Delta$ for all i). Then for $p = 1$, $\tilde{\alpha}_0 = -\frac{1}{\Delta}\left[\frac{\hat{\gamma}(\Delta)}{\hat{\gamma}(0)} - 1\right]$ where $\hat{\gamma}(\Delta) = \frac{1}{N}\sum_{i=1}^{N-1} x(t_i)x(t_{i+1})$ and

$$E(\tilde{\alpha}_0) = -\frac{1}{\Delta}\left[e^{-\alpha_0\Delta} - 1\right] + O(N^{-1}).$$

Thus the discrete time approximation to the Yule–Walker estimator is asymptotically unbiased (as $N \to \infty$ and $\Delta \to 0$).

But for $p > 1$, (3.1.12) gives an asymptotically biased estimate of $D_{j,p}$. This is most easily seen by considering the case $p = 2$ with $\hat{x}^{(1)}(t_i) := [x(t_{i+1}) - x(t_i)]/\Delta$. (A similar analysis to that below is given by Bartlett (1946, p.39).) By the mean value theorem, $\hat{x}^{(1)}(t_i) = x^{(1)}(\xi)$ for some $\xi \in (t_i, t_{i+1})$. Thus, the integrand in $\tilde{D}_{1,2}$ is not evaluated at t_i and so $\tilde{D}_{1,2}$ does not converge to the

required Itô integral. Specifically,

$$\begin{aligned} E(\tilde{D}_{1,2}) &= E\left\{\frac{1}{T}\sum_i \hat{x}^{(1)}(t_i)\left[\hat{x}^{(1)}(t_{i+1}) - \hat{x}^{(1)}(t_i)\right]\right\} \\ &= E\Bigg\{\frac{1}{N\Delta^3}\sum_i \Big[x(t_{i+1})x(t_{i+2}) - x(t_i)x(t_{i+2}) - 2x^2(t_{i+1}) \\ &\qquad + 2x(t_i)x(t_{i+1}) + x(t_{i+1})x(t_i) - x^2(t_i)\Big]\Bigg\} \\ &\approx \frac{1}{\Delta^3}\left\{-3\gamma(0) + 4\gamma(\Delta) - \gamma(2\Delta)\right\} \quad \text{for large } N. \end{aligned} \tag{3.1.13}$$

Now $\gamma(h)$ is the top left element of $\Gamma_p(h) = e^{Ah}\Gamma_p(0)$ using (2.3.9). For small h, $e^{Ah} \approx I + Ah + A^2h^2/2 + A^3h^3/6$. So the top left element of e^{Ah} is approximately $1 - \alpha_0 h^2/2 + \alpha_0\alpha_1 h^3/6$ and

$$\gamma(h) \approx \gamma(0)\left[1 - \tfrac{\alpha_0 h^2}{2} + \tfrac{\alpha_0\alpha_1 h^3}{6}\right]. \tag{3.1.14}$$

Substituting this expression into (3.1.13) gives $E(\tilde{D}_{1,2}) \approx -2\alpha_0\alpha_1\gamma(0)/3$ for small Δ and large N. Similarly, $E(\tilde{D}_{1,1}) \approx \gamma(0)\alpha_0(1 - \alpha_1\Delta/3)$. But using (2.3.7) and the results of Section 3.1.4, $D_{1,2} = -\alpha_0\alpha_1\gamma(0) \approx \frac{3}{2}E(\tilde{D}_{1,2})$. Hence $E(\tilde{\alpha}_1) = E(-\tilde{D}_{1,2}/\tilde{D}_{1,1}) \approx 2\alpha_1/3$ for large N and small Δ. The estimate of α_0 is not asymptotically biased in this way since it does not rely on $\tilde{D}_{1,2}$. The above approximations give $E(\tilde{\alpha}_0) = E(\tilde{D}_{1,1}/\tilde{D}_{0,0}) \approx \alpha_0 - \alpha_0\alpha_1\Delta/3$.

Using other methods of differencing does not resolve the problem. For instance, with central differences, $\hat{x}^{(1)}(t_i) := [x(t_{i+1}) - x(t_{i-1})]/2\Delta$. Then, by the mean value theorem, $\hat{x}^{(1)}(t_i) = x^{(1)}(\xi)$ for some $\xi \in (t_{i-1}, t_{i+1})$. So, again, the integrand in $\tilde{D}_{1,2}$ is not evaluated at t_i and so $\tilde{D}_{1,2}$ does not converge to the

required Itô integral. In this case,

$$E(\tilde{D}_{1,2}) \approx \frac{1}{4\Delta^3}\{-2\gamma(0) + \gamma(\Delta) + 2\gamma(2\Delta) - \gamma(3\Delta)\} \quad \text{for large } N.$$

Then, substituting (3.1.14) gives $E(\tilde{D}_{1,2}) \approx -5\alpha_0\alpha_1\gamma(0)/12$. Similarly, $E(\tilde{D}_{1,1}) \approx \gamma(0)\alpha_0(1 - 2\alpha_1\Delta/3)$. Hence $E(\tilde{\alpha}_1) \approx 5\alpha_1/12$ and $E(\tilde{\alpha}_0) \approx \alpha_0 - 2\alpha_0\alpha_1\Delta/3$. Thus, using central differences results in even more bias than forward differences!

Clearly, for discrete time data, another approach is required. If the data are equally spaced, one approach is to equate the ACVF with its sample estimate rather than equating the DCVF with its sample estimate. This continues the flavour of Yule–Walker estimation for discrete autoregressive models although the resulting estimates can no longer be calculated by solving a set of p linear equations. In fact, we obtain a set of p non-linear equations which can be solved with only a little more computation.

A modification of this approach suitable for unequally spaced data will be given at the end of this section.

Equally spaced data

First, consider the simple CAR(1) model observed at equally spaced intervals of length Δ. Then

$$\gamma(\Delta) = e^{-\alpha_0\Delta}\gamma(0).$$

This suggests the estimator

$$\tilde{\tilde{\alpha}}_0 := -\frac{1}{\Delta}\log\left[\frac{\hat{\gamma}(\Delta)}{\hat{\gamma}(0)}\right] = -\frac{1}{\Delta}\log\left[1 - \Delta\tilde{\alpha}_0\right].$$

Now the restriction of small Δ can be dropped. Note that $e^{-\tilde{\tilde{\alpha}}_0\Delta}$ is the discrete time Yule–Walker estimator for the discrete time AR(1) process formed by the observations. Of course, $\tilde{\tilde{\alpha}}_0$ is also biased because of the ratio estimate of $\gamma(\Delta)/\gamma(0)$. The estimate may be further improved by replacing $\hat{\gamma}(\Delta)/\hat{\gamma}(0)$ with a jackknife estimator.

In general, we have a set of p non-linear equations

$$f_k(\boldsymbol{\alpha}) = \gamma(k\Delta) - \hat{\gamma}(k\Delta), \qquad k = 1, 2, \ldots, p.$$

We wish to find $\boldsymbol{\alpha}$ such that $f_k(\boldsymbol{\alpha}) = 0$ for all k. For $p > 1$ a closed form solution is not possible although the equations may be solved using the Newton–Raphson method as follows. Expanding f_k in a Taylor series yields

$$f_k(\boldsymbol{\alpha} + \delta\boldsymbol{\alpha}) = f_k(\boldsymbol{\alpha}) + \sum_{j=0}^{p-1} \frac{\partial f_k}{\partial \alpha_j}\delta\alpha_j + O(\delta\boldsymbol{\alpha}^2). \tag{3.1.15}$$

Ignoring the terms of order $\delta\boldsymbol{\alpha}^2$ and higher, we may calculate the corrections $\delta\boldsymbol{\alpha}$ from the set of linear equations

$$\sum_{j=0}^{p-1} f_{k,j}\delta\alpha_j = \beta_k \tag{3.1.16}$$

where $f_{k,j} = \frac{\partial f_k}{\partial \alpha_j}$ and $\beta_k = -f_k$. Given initial values, $\boldsymbol{\alpha}$, equation (3.1.15) may be iterated to convergence.

To calculate f_k and $f_{k,j}$ we need to find $\gamma(k\Delta)$ given $\boldsymbol{\alpha}$. This may be determined from the top left element of $e^{Ak\Delta}\Gamma_p(0)$. Elements of e^{Ah} may be calculated using Taylor series approximations and elements of $\Gamma_p(0)$ may be estimated using $\tilde{D}_{j,k}$.

Thus, we obtain

$$\gamma(h) \approx \sum_{k=0}^{p-1} \hat{\gamma}^{(k)}(0) \sum_{m=0}^{M} \frac{c_{m,k} h^m}{m!} \tag{3.1.17}$$

for some sufficiently large M where $\hat{\gamma}^{(2j)}(0) = \tilde{D}_{j,j}$, $\hat{\gamma}^{(2j-1)}(0) = 0$, $2j \leq p$, and $c_{m,k}$ is a function of the coefficients $\alpha_0, \alpha_1, \ldots, \alpha_{p-1}$.

For example, if $p = 2$, further terms may be added to (3.1.14) to give

$$\gamma(h) \approx \hat{\gamma}(0) \Big[1 - \tfrac{\alpha_0 h^2}{2} + \tfrac{\alpha_0 \alpha_1 h^3}{6} + \tfrac{\alpha_0(-\alpha_1^2+\alpha_0)h^4}{24} + \tfrac{\alpha_0(\alpha_1^3-2\alpha_0\alpha_1)h^5}{120} + \tfrac{\alpha_0(-\alpha_1^4+3\alpha_0\alpha_1^2-\alpha_0^2)h^6}{720}\Big]. \tag{3.1.18}$$

Then f_k and $f_{k,j}$ may be easily calculated for all k and j. In particular,

$$f_{k,0}(\boldsymbol{\alpha}) \approx \gamma(0) \Big[-\tfrac{h^2}{2} + \tfrac{\alpha_1 h^3}{6} + \tfrac{(-\alpha_1^2+2\alpha_0)h^4}{24} + \tfrac{(\alpha_1^3-4\alpha_0\alpha_1)h^5}{120} + \tfrac{(-\alpha_1^4+6\alpha_0\alpha_1^2-3\alpha_0^2)h^6}{720}\Big]$$

and $$f_{k,1}(\boldsymbol{\alpha}) \approx \gamma(0) \Big[\tfrac{\alpha_0 h^3}{6} - \tfrac{2\alpha_0\alpha_1 h^4}{24} + \tfrac{\alpha_0(3\alpha_1^2-2\alpha_0)h^5}{120} + \tfrac{\alpha_0(-4\alpha_1^3+6\alpha_0\alpha_1)h^6}{720}\Big]$$

where $h = k\Delta$. Again, we do not need Δ small for this approximation, although for $p > 2$ this restriction is necessary to calculate $\hat{\gamma}^{(2j)}(0)$, $j \geq 1$.

For $p = 3$ and $p = 4$, values of $c_{m,k}$ are given in Table 1.

For equally spaced CAR(2) processes, Bartlett (1946, 1955) suggested a similar approach identifying the sample and theoretical autocovariances, but he calculated the theoretical covariances using the relation

$$\gamma(h) = \frac{e^{-\frac{1}{2}\alpha_1|h|}\cos(\lambda|h| - \theta)}{\cos\theta}\gamma(0)$$

where $\lambda = \sqrt{\alpha_0 - \alpha_1^2/4}$, $\tan\theta = \alpha_1/(2\lambda)$ and $\alpha_0 \geq \alpha_1^2/4$.

	$p=1$	$p=2$	$p=3$	$p=4$
$c_{0,0}$	1	1	1	1
$c_{1,0}$	$-\alpha_0$	0	0	0
$c_{2,0}$	α_0^2	$-\alpha_0$	0	0
$c_{3,0}$	$-\alpha_0^3$	$\alpha_0\alpha_1$	$-\alpha_0$	0
$c_{4,0}$	α_0^4	$\alpha_0(-\alpha_1^2+\alpha_0)$	$\alpha_0\alpha_2$	$-\alpha_0$
$c_{5,0}$	$-\alpha_0^5$	$\alpha_0(\alpha_1^3-2\alpha_0\alpha_1)$	$\alpha_0(-\alpha_2^2+\alpha_1)$	$\alpha_0\alpha_3$
$c_{6,0}$	α_0^6	$\alpha_0(-\alpha_1^4+3\alpha_0\alpha_1^2-\alpha_0^2)$	$\alpha_0(\alpha_2^3-2\alpha_1\alpha_2+\alpha_0)$	$\alpha_0(-\alpha_3^2+\alpha_2)$
$c_{7,0}$	$-\alpha_0^7$	$\alpha_0(\alpha_1^5-4\alpha_0\alpha_1^3+3\alpha_0^2\alpha_1)$	$\alpha_0(-\alpha_2^4+3\alpha_1\alpha_2^2-2\alpha_0\alpha_2+\alpha_1^2)$	$\alpha_0(\alpha_3^3-2\alpha_2\alpha_3+\alpha_1)$
$c_{0,2}$			0	0
$c_{1,2}$			0	0
$c_{2,2}$			1	1
$c_{3,2}$			$-\alpha_2$	0
$c_{4,2}$			$\alpha_2^2-\alpha_1$	$-\alpha_2$
$c_{5,2}$			$-\alpha_2^3+2\alpha_1\alpha_2-\alpha_0$	$\alpha_2\alpha_3-\alpha_1$
$c_{6,2}$			$\alpha_2^4-3\alpha_1\alpha_2^2+2\alpha_0\alpha_2+\alpha_1^2$	$-\alpha_2\alpha_3^2+\alpha_1\alpha_3-\alpha_0+\alpha_2^2$
$c_{7,2}$			$-\alpha_2^5+4\alpha_1\alpha_2^3-3\alpha_0\alpha_2^2-3\alpha_1^2\alpha_2+2\alpha_0\alpha_1$	$\alpha_2\alpha_3^3-\alpha_1\alpha_3^2-2\alpha_2^2\alpha_3+2\alpha_1\alpha_2+\alpha_0\alpha_3$

Table 1: *Values of $c_{m,k}$ for use with equation (3.1.17).*

Unequally spaced data

Let $\bar{\gamma}_k = \frac{1}{N}\sum_i X(t_{i+k})X(t_i)$ so that $E[\bar{\gamma}_k] = \frac{1}{N}\sum_i \gamma(\Delta_{i,k})$ where $\Delta_{i,k} = t_{i+k} - t_i$. Note that for equally spaced data, $\bar{\gamma}_k = \hat{\gamma}(k\Delta)$. Using (3.1.17) and interchanging the order of summation, we obtain

$$E[\bar{\gamma}_k] \approx \sum_{k=0}^{p-1} \hat{\gamma}^{(k)}(0) \sum_{m=0}^{M} \frac{c_{m,k}}{m!} \frac{1}{N} \sum_{i=1}^{N-k} \Delta_{i,k}^m.$$

Hence, the estimates derived above may be modified for unequally spaced data by replacing $\hat{\gamma}(k\Delta)$ with $\bar{\gamma}_k$, $\gamma(k\Delta)$ with $E[\bar{\gamma}_k]$ and $(k\Delta)^m$ with $\frac{1}{N}\sum_i \Delta_{i,k}^m$.

3.2 Maximum Likelihood Estimation

Suppose $\{X(t)\}$ is observed at the times $t_1 < t_2 < \cdots < t_N$ where these times may be unequally spaced. Let $X(t_i \,|\, t_j) = E[X(t_i) \,|\, X(t_k) = x(t_k), k \leq j]$ and denote the innovation (or one-step-ahead prediction error) by $I(t_i) = X(t_i) - X(t_i \,|\, t_{i-1})$ and its variance by $V(t_i) = E[I^2(t_i) \,|\, X(t_k) = x(t_k), k < i]$.

Then, the likelihood of the observations $x(t_1), \ldots, x(t_N)$ is

$$L = (2\pi)^{-N/2}[V(t_1)\ldots V(t_N)]^{-1/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{N} I(t_i)^2/V(t_i)\right],$$

where $X(t_1 \,|\, t_0) = \frac{-\beta}{\alpha_0}$ and $V(t_1)$ is the top left component of the stationary state covariance matrix, $\Gamma_p(0) = \sigma^2 \int_0^\infty e^{Ay}\boldsymbol{b}\boldsymbol{b}^T e^{A^T y} dy$.

A non-linear optimisation algorithm can then be used in conjunction with the expression for L to find maximum likelihood estimates of the parameters.

Equivalently, one can minimise

$$-2\log L = N\log(2\pi) + \sum_{i=1}^{N} \log V(t_i) + \sum_{i=1}^{N} I^2(t_i)/V(t_i). \tag{3.2.1}$$

For $p = 1$, $X(t_i \,|\, t_{i-1}) = m(x(t_{i-1}), t_i - t_{i-1})$ and $V(t_i) = v(x(t_{i-1}), t_i - t_{i-1})$ where

$$m(x,t) = e^{-\alpha_0 t}x - \frac{\beta}{\alpha_0}\left[1 - e^{-\alpha_0 t}\right], \tag{3.2.2}$$

$$\text{and} \quad v(x,t) = \frac{\sigma^2}{2\alpha_0}\left[1 - e^{-2\alpha_0 t}\right] \tag{3.2.3}$$

as defined in Section 2.1.1.

For higher order processes, $I(t_i)$ and $V(t_i)$ may be obtained from the Kalman recursions.

3.2.1 Kalman recursions

Consider the state space form of the CAR(p) process,

$$\begin{aligned} d\boldsymbol{S}_*(t) &= A\boldsymbol{S}_*(t)dt + \sigma \boldsymbol{b} dW(t) \\ X(t) &= [1, 0, \cdots, 0]\boldsymbol{S}_*(t) \end{aligned}$$

where

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -\alpha_0 & -\alpha_1 & \cdots & \cdots & -\alpha_{p-1} \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

$$\text{and} \qquad \boldsymbol{S}_*(t) = \begin{bmatrix} X(t) + \frac{\beta}{\alpha_0} \\ X^{(1)}(t) \\ \vdots \\ X^{(p-2)}(t) \\ X^{(p-1)}(t) \end{bmatrix}.$$

From (2.1.5), we obtain the solution to the state space equation at each observation

$$\begin{aligned} \boldsymbol{S}_*(t_{i+1}) &= e^{At}\boldsymbol{S}_*(0) + \sigma \int_0^{t_{i+1}} \exp[A(t_{i+1} - y)]\boldsymbol{b}dW(y) \\ &= \exp[A(t_{i+1} - t_i)]\boldsymbol{S}_*(t_i) + \sigma \int_{t_i}^{t_{i+1}} \exp[A(t_{i+1} - y)]\boldsymbol{b}dW(y). \end{aligned}$$

Hence, the system can be written as a state space model in discrete time

$$\begin{aligned} \boldsymbol{S}_*(t_{i+1}) &= F_i\boldsymbol{S}_*(t_i) + \boldsymbol{G}_i \\ X(t_i) &= \boldsymbol{a}^T\boldsymbol{S}_*(t_i) \end{aligned}$$

where

$$\begin{aligned} F_i &= \exp[A(t_{i+1} - t_i)] \\ \boldsymbol{G}_i &= \sigma \int_{t_i}^{t_{i+1}} \exp[A(t_{i+1} - y)]\boldsymbol{b}dW(y) \\ \text{and} \qquad \boldsymbol{a} &= [1, 0, \cdots, 0]^T. \end{aligned}$$

Thus $\{\boldsymbol{G}_i\}$ is an independent sequence of Gaussian random vectors with mean $E[\boldsymbol{G}_i] = \boldsymbol{0}$ and covariance matrix

$$E[\boldsymbol{G}_i\boldsymbol{G}_i^T] = \sigma^2 \int_0^{t_{i+1}-t_i} e^{Ay}\boldsymbol{b}\boldsymbol{b}^T e^{A^T y}dy.$$

These equations are in precisely the form needed for application of the Kalman recursions (see, for example, Brockwell and Davis, 1991, Chapter 12).

However, it is difficult to compute the matrix exponential numerically (see Moler and van Loan, 1978). Jones (1981) discusses a solution to this difficulty by using the spectral representation of the matrix A.

If it is assumed that the eigenvalues of A are distinct, then $A = U\Lambda U^{-1}$ where Λ is a diagonal matrix of the eigenvalues of A and U is a $p \times p$ matrix, the columns of which are the right eigenvectors of A. Thus,

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_p \end{pmatrix} \quad \text{and} \quad U = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_p \\ \vdots & \vdots & & \vdots \\ \lambda_1^{p-1} & \lambda_2^{p-1} & \cdots & \lambda_p^{p-1} \end{pmatrix}$$

where λ_k, $k = 1, \ldots, p$, is a root of $\mathcal{C}(z) = \sum_{j=0}^{p} \alpha_p z^j = 0$.

Then, defining a transformed state vector $\boldsymbol{Y}(t) = U^{-1}\boldsymbol{S}_*(t)$, the new state space equations are

$$\begin{aligned} d\boldsymbol{Y}(t) &= \Lambda \boldsymbol{Y}(t)dt + \sigma \boldsymbol{C} dW(t) \\ \text{and} \qquad X_*(t) &= D\boldsymbol{Y}(t) \end{aligned}$$

where $\boldsymbol{C} = U^{-1}\boldsymbol{b}$ and $D = \boldsymbol{a}^T U = [1, 1, \cdots, 1]$. Likewise, the observations satisfy the discrete time state space equations

$$\begin{aligned} \boldsymbol{Y}(t_{i+1}) &= B_i \boldsymbol{Y}(t_i) + \boldsymbol{w}_i \\ X_*(t_i) &= D\boldsymbol{Y}(t_i) \end{aligned}$$

where

$$B_i = \exp[\Lambda(t_{i+1} - t_i)]$$

$$\text{and} \quad \boldsymbol{w}_i = \sigma \int_{t_i}^{t_{i+1}} \exp[\Lambda(t_{i+1} - y)]\boldsymbol{C}dW(y).$$

The advantage of these transformed equations is that the matrix exponential, B_i can be calculated as the diagonal matrix with entries $\exp[\lambda_k(t_{i+1}-t_i)]$ where λ_k is an eigenvalue of A.

The covariance matrix of $\boldsymbol{w}_i$ is

$$Q_i = \sigma^2 \int_{t_i}^{t_{i+1}} \exp[\Lambda(t_{i+1} - y)]\boldsymbol{C}\bar{\boldsymbol{C}}^T \exp[\bar{\Lambda}^T(t_{i+1} - y)]dy$$

where $\bar{M}$ denotes the complex conjugate of the matrix M. Thus, the (j,k)th component of Q_i is

$$[Q_i]_{j,k} = \begin{cases} \frac{\sigma^2 c_j \bar{c}_k}{\lambda_j + \bar{\lambda}_k}\left\{\exp[(\lambda_j + \bar{\lambda}_k)(t_{i+1} - t_i)] - 1\right\} & \lambda_j + \bar{\lambda}_k \neq 0 \\ \sigma^2 c_j \bar{c}_k (t_{i+1} - t_i) & \lambda_j + \bar{\lambda}_k = 0 \end{cases}$$

where c_j denotes the jth component of $\boldsymbol{C}$.

Now let $\boldsymbol{Y}(t_i \,|\, t_j)$, denote the estimate of $\boldsymbol{Y}(t_i)$ given the observations up to and including time t_j and let $P(t_i \,|\, t_j)$ denote the covariance of $\boldsymbol{Y}(t_i \,|\, t_j)$.

Then the Kalman filter algorithm allows iterative calculation of these quantities:

$$\boldsymbol{Y}(t_{i+1} \,|\, t_i) = B_i \boldsymbol{Y}(t_i \,|\, t_i) \tag{3.2.4}$$

$$P(t_{i+1} \,|\, t_i) = B_i P(t_i \,|\, t_i)\bar{B}_i^T + Q_i \tag{3.2.5}$$

$$X_*(t_{i+1} \,|\, t_i) = D\boldsymbol{Y}(t_{i+1} \,|\, t_i) \tag{3.2.6}$$

$$I(t_{i+1}) = X_*(t_{i+1}) - X_*(t_{i+1} \,|\, t_i) \tag{3.2.7}$$

$$V(t_{i+1}) = DP(t_{i+1} \,|\, t_i)\bar{D}^T \tag{3.2.8}$$

$$\boldsymbol{Y}(t_{i+1} \,|\, t_{i+1}) = \boldsymbol{Y}(t_{i+1} \,|\, t_i) + P(t_{i+1} \,|\, t_i)\bar{D}^T I(t_{i+1})/V(t_{i+1}) \tag{3.2.9}$$

$$P(t_{i+1} \,|\, t_{i+1}) = P(t_{i+1} \,|\, t_i) - P(t_{i+1} \,|\, t_i)\bar{D}^T DP(t_{i+1} \,|\, t_i)/V(t_{i+1}). \tag{3.2.10}$$

3.3 Modelling Paradigm

This section outlines a modelling paradigm for data observed at deterministic, possibly irregularly spaced, times. This involves data transformations, model estimation, order selection, diagnostic checking and forecasting. Aspects of this approach may be found in Jones (1981, 1984).

3.3.1 Time scale

Suppose a CAR process is observed at times Δ apart. Then a change in the units in which Δ is expressed results in a change in the model parameters. To prevent very small or very large model parameters, it is necessary to choose the time scale such that Δ is not too large or too small.

For example, consider the CAR(1) process

$$\frac{dX(t)}{dt} + \alpha_0 X(t) = \sigma Z(t).$$

Now $\frac{dX(t)}{dt}$ will double if the time scale is halved whereas $X(t)$ will stay the same. Hence, α_0 will be doubled. That is, α_0 is inversely proportional to the time unit.

Generally, for a CAR(p) process, $X^{(j)}(t)$, $j < p$, is inversely proportional to the jth power of the time unit. Therefore, α_j is inversely proportional to the $(p-j)$th power of the time unit. Also the variance of $\{X(t)\}$ is invariant on different time scales. Now, the roots of the equation will be proportional to the time scale. Thus, from (3.1.9) and (3.1.10), σ^2 will be inversely proportional to the $(2p-1)$th root of the time unit.

So, a CAR(2) process with parameters $\beta = 0$, $\alpha_0 = 0.5$, $\alpha_1 = 0.5$ and $\sigma^2 = 1.0$ and time unit Δ is identical to the CAR(2) process with parameters $\beta = 0$, $\alpha_0 = 2.0$, $\alpha_1 = 1.0$ and $\sigma^2 = 8.0$ defined on a time scale with time unit $\Delta/2$.

Clearly, the numerical computations are subject to excessive rounding error if the spacing between observations is too small or too great. The above discussion shows how to change the time scale appropriately when necessary.

For example, Jones (1984) fits a CAR(1) model to data relating to the DNA binding in a patient with systemic lupus erythematosis. There are 99 observations over a period of 3061 days. The shortest time between observations was 1 day, the largest was 90 days and the average 31.2 days. Using the natural time unit of one day, $\hat{\alpha}_0 = 0.0077$ and $\hat{\sigma}^2 = 0.0980$. By altering the time scale so that 1 time unit = 30 days, we obtain $\hat{\alpha}_0 = 30 \times 0.0077 = 0.231$ and $\hat{\sigma}^2 = 30 \times 0.0980 = 2.94$. Jones uses a time unit of 10 days which gives $\hat{\alpha}_0 = 10 \times 0.0077 = 0.077$ and $\hat{\sigma}^2 = 10 \times 0.0980 = 0.980$.

3.3.2 Data transformations

In estimating the model, it is assumed that the data are stationary and normally distributed. Graphical methods are usually sufficient to check these

assumptions. Where the data are clearly non-stationary or non-Gaussian, Box–Cox transformations or power transformations may be used.

The algorithms used for inference are more stable if the data do not have a very low coefficient of variation. To avoid this problem, it may be necessary to rescale the data. Jones (1984) normalises the data by subtracting the sample mean and dividing by the sample standard deviation. Yeung (1989) also recommends this procedure.

3.3.3 Order selection criteria

The AIC (Akaike Information Criterion) statistic was introduced by Akaike (1969) and may be defined as

$$AIC := -2\log L + 2k$$

where L denotes the likelihood of the model and k is the number of parameters estimated in the model. Thus the AIC penalises the log likelihood by k. For a CAR(p) model with non-zero mean, $k = p + 1$. Minimising the AIC statistic provides a useful criterion for choosing the model order.

An alternative approach is the BIC (Bayesian Information Criterion) which is identical to the AIC except the penalty is $\frac{k}{2}\log N$. That is,

$$BIC := -2\log L + k\log N.$$

This was introduced by Akaike (1977) to correct the tendency of the AIC to overestimate the number of parameters.

I have found the Yule–Walker estimates to be adequate in assessing model order. For each order the parameters are estimated using the Yule–Walker approach described in Section 3.1.5 and the above statistics are calculated. If the Yule–Walker estimates do not converge or give a non-stationary model, this may be interpreted to mean that a model of that order is not appropriate for the data.

To test the effectiveness of this approach to order selection, a simulation study was carried out where CAR(1) processes were simulated and the order selection criteria applied. One thousand samples were generated for each value of the parameters. The processes had parameters $\sigma = 1$, $\beta = 0$ and α_0 was either 0.1, 0.5, 1.0, 2.0 or 3.0. Each simulated series consisted of $N = 50$ observations at unit intervals. The simulations were repeated for $N = 100$ and $N = 1000$. The results are given in Table 2.

	$\boldsymbol{\alpha_0}$	AIC: 1 BIC: 1	AIC: 2 BIC: 1	AIC: 2 BIC: 2	AIC: 3 BIC: 3	**Total**
$N = 50$	0.1	957	18	22	0	997
	0.5	928	40	32	0	1000
	1.0	889	50	57	0	996
	2.0	638	198	63	3	902
	3.0	437	52	290	2	781
$N = 100$	0.1	985	5	9	0	999
	0.5	979	9	12	0	1000
	1.0	966	15	19	0	1000
	2.0	753	57	154	1	965
	3.0	429	24	336	4	793
$N = 1000$	0.1	1000	0	0	0	1000
	0.5	1000	0	0	0	1000
	1.0	1000	0	0	0	1000
	2.0	1000	0	0	0	1000
	3.0	860	1	138	0	999

Table 2: *Results of simulations of CAR(1) processes*

Cases where the Yule–Walker estimates did not converge or gave a non-

stationary model have been dropped from the above table (hence the figures less than 1000 in the right hand column). These results confirm that the AIC tends to overestimate the model order, particularly for large values of α_0 and small values of N. Note that in these simulations, it is not possible for the BIC to select order 2 when the AIC selects order 1 since $N > e^2$.

A similar analysis was carried out for CAR(2) processes with $\sigma = 1$, $\beta = 0$ and α_0 and α_1 taking the values 0.1, 0.5, 1.0 and 2.0. Again, one thousand series with unit observation spacing were generated. Here $N = 100$. The results are given in Table 3.

$\boldsymbol{\alpha_0}$	$\boldsymbol{\alpha_1}$	AIC: 1 BIC: 1	AIC: 2 BIC: 1	AIC: 2 BIC: 2	AIC: 3 BIC: 2	AIC: 3 BIC: 3	**Total**
0.1	0.1	485	0	515	0	0	1000
0.1	0.5	147	2	848	1	1	999
0.1	1.0	91	5	897	4	3	1000
0.1	2.0	469	83	437	0	3	992
0.5	0.1	194	0	803	0	3	1000
0.5	0.5	0	0	995	5	0	1000
0.5	1.0	7	0	993	0	0	1000
0.5	2.0	370	107	523	0	0	1000
1.0	0.1						
1.0	0.5	0	0	1000	0	0	1000
1.0	1.0	0	0	1000	0	0	1000
1.0	2.0	238	116	646	0	0	1000
2.0	0.1						
2.0	0.5	0	0	999	0	1	1000
2.0	1.0	0	0	996	3	1	1000
2.0	2.0	274	155	571	0	0	1000

Table 3: *Results of simulations of CAR(2) processes*

Again, cases where the Yule–Walker estimates did not converge or gave a non-stationary model have been dropped from the above table (hence the figures less than 1000 in the right hand column). This only occurred when α_0/α_1 was small. On the other hand, when α_0/α_1 was large (> 10), the simulated series was unstable since the process is close to the boundary of stationarity.

Therefore no results were recorded for these processes.

The results indicate that the BIC often underestimates the model order for CAR(2) processes while the AIC rarely overestimates the order. The best results are obtained for moderate values of α_0 and α_1.

This brief Monte-Carlo study indicates that both the AIC and BIC should be used. Where the results differ, the choice between models is best made on other grounds such as the behaviour of the residuals.

3.3.4 Model estimation

The discrete time Yule–Walker estimates described in Section 3.1.5 can be used as starting values for maximum likelihood estimation.

3.3.5 Diagnostics

If the data are from a CAR process, the innovations, $I(t_i)$, are Gaussian and orthogonal to past observations. So, the standardised one-step-ahead prediction errors, given by $e(t_i) = I(t_i)/\sqrt{V(t_i)}$, are i.i.d. standard normal random variates and each error is orthogonal to all observations taken at times prior to t_i. Thus they provide a useful approach to testing the goodness-of-fit of a CAR model. For example, tests for normality and serial correlation may be applied as for discrete time models (see Brockwell and Davis, 1991, Section 9.4).

3.3.6 Forecasting

To calculate the minimum mean square predictor of $X(t_h)$ and its variance, one simply needs to calculate $X(t_h|t_N)$ and $V(t_h)$, $h > N$. For $p = 1$, this may be accomplished using $m(x(t_N), t_h - t_N)$ and $v(x(t_N), t_h - t_N)$ respectively. More generally, the Kalman recursions may be used. After running through equations (3.2.4) – (3.2.10), $X(t_h|t_N)$ and $V(t_h)$ may be calculated using equations (3.2.4) – (3.2.6) and (3.2.8).

The prediction density is Gaussian and so prediction intervals may be obtained using these first two moments.

3.4 Applications

In this section, the modified Yule–Walker estimators derived in Section 3.1 and the maximum likelihood estimators described in Section 3.2 are applied to three data sets which have been subject to previous analyses.

3.4.1 IBM Closing Stock Prices

The IBM daily closing stock price data, $\{P(t)\}$, are given in Box and Jenkins (1976, p.526). I shall consider the modelling of the relative daily price changes, $X(t) = 100[P(t) - P(t-1)]/P(t-1)$ for the period 18 May 1961 – 30 March 1962. This series is also considered by Yeung (1989) and Tong and Yeung (1991). I shall treat the data as uniformly spaced, although it would not be difficult to treat the data as irregularly spaced to take into account weekends

and holidays (Tong and Yeung treat the data both ways). The relative price changes are plotted in Figure 3.4.1.

Figure 3.4.1: *Relative changes of IBM daily closing stock prices (18 May 1961 – 30 March 1962).*

The histogram of the data is approximately symmetric and so no transformation was used. Both the AIC and BIC indicate the use of a CAR(1) model in accordance with the analysis of Tong and Yeung.

The Yule–Walker model is

$$X^{(1)}(t) + 1.563X(t) - 0.120 = 1.695W^{(1)}(t).$$

The one-step MSE of this model is 0.879 and the log-likelihood is -295.327. Using this model as a starting point for maximum likelihood estimation, we obtain

$$X^{(1)}(t) + 1.559X(t) - 0.109 = 1.693W^{(1)}(t)$$

where the parameters β, α_0 and σ have standard errors 0.125, 0.288 and 0.154 respectively and the model gives one-step MSE 0.879 and log-likelihood -295.324.

Clearly, the Yule–Walker method gives an excellent starting point very close to the maximum likelihood result.

Yeung (1989) and Tong and Yeung (1991) also use maximum likelihood estimation and give the estimates $\hat{\alpha}_0 = 1.561$ and $\hat{\sigma} = 1.764$.

3.4.2 Wolf's Sunspot Numbers

The annual means of the annual Wolf sunspot numbers for 1700–1988 are given in Tong (1990, p.470). These are plotted in Figure 3.4.2. This series has been

Figure 3.4.2: *Annual means of Wolf's sunspot numbers (1700–1988).*

analysed many times in the statistical literature. In fact, autoregressive models were first introduced for the sunspot numbers by Yule (1927).

The histogram of the data suggest a power transformation is necessary. After taking square roots of the series, the histogram of the transformed data is roughly symmetrical. This choice of transformation has also been used by several authors including Jones (1981), Ghaddar and Tong (1981), Chan and Tong (1986) and Yeung (1989). However, other authors model the raw sunspot numbers.

Using either the raw data or the transformed data, the AIC and BIC indicate a second order model. This is in accordance with most of the autoregressive models of these data that have appeared in the literature, starting with Yule's original analysis in 1927.

Various subsets of the series have been analysed. Several authors have considered CAR(2) models for the raw sunspot numbers for the period 1749–1924. Their parameter estimates are given below.

Reference	$\boldsymbol{\alpha_0}$	$\boldsymbol{\alpha_1}$
Bartlett (1946)	0.616	0.316
Phadke and Wu (1974) A	0.479	0.632
Phadke and Wu (1974) B	0.503	0.785
Robinson (1980)	0.511	0.793
Yeung (1989)	0.433	0.982

Bartlett (1946) used a method equivalent to the Yule–Walker method given in Section 3.1, by equating the sample autocorrelations with the theoretical model autocorrelations. However, his results seem incorrect. Phadke and Wu (1974)

recalculated the parameters using Bartlett's method and obtained results (denoted A above) very close to those obtained using my own implementation of the Yule–Walker approach with $M = 7$ in equation (3.1.17).

The discrete time approximation of the Yule–Walker estimators for these data gives a CAR(2) model with parameters $\tilde{\alpha}_0 = 0.376$ and $\tilde{\alpha}_1 = 0.510$. Starting with these values and using $M = 7$, (3.1.15) was iterated to obtain the model

$$X^{(2)}(t) + 0.617X^{(1)}(t) + 0.478X(t) - 21.480 = 26.654W^{(1)}(t)$$

with one-step MSE 263.5 and log-likelihood -740.1.

Phadke and Wu (1974) also calculated the parameters by estimating a discrete time ARMA(2,1) model using maximum likelihood and calculating the corresponding CAR(2) model. The results are denoted B above. This approach is equivalent to the direct estimation of the CAR(2) parameters using maximum likelihood estimation, although the method is not easily generalised to irregularly spaced data. Robinson (1980) also used an efficient method and, not surprisingly, obtained very similar results to Phadke and Wu (1974) B.

My own implementation of the maximum likelihood approach yields the model

$$X^{(2)}(t) + 0.789X^{(1)}(t) + 0.500X(t) - 22.420 = 30.633W^{(1)}(t)$$

where the parameters β, α_0, α_1 and σ have standard errors 0.970, 0.042, 0.088 and 0.197 respectively and the model gives one-step MSE 261.1 and log-likelihood -738.4.

Yeung (1989) also claims to use the maximum likelihood method described in Section 3.2. However, a comparison with the above equation and the results

of Phadke and Wu (1974) B and Robinson (1980) indicates that she may have made some calculation error.

3.4.3 Canadian Lynx Data

This series consists of the numbers of Canadian lynx trapped annually in the Mackenzie River district of north-west Canada for the period 1821–1934. The series is given in Tong (1990, p.469) and shown in Figure 3.4.3.

Figure 3.4.3: *Numbers of Canadian lynx trapped annually in the Mackenzie River district of north-west Canada (1821–1934).*

The Canadian lynx data is a classic series in the time series literature and has received extensive attention. Examination of the histogram shows that a transformation is required. The log (base 10) transformation seems to give a roughly symmetric histogram. This choice has been favoured in most of the analyses of the data since the first time series study of the data was undertaken

by Moran (1953).

The discrete time approximation of the Yule–Walker estimators for these data gives a CAR(2) model with parameters $\tilde{\alpha}_0 = 0.417$ and $\tilde{\alpha}_1 = 0.438$. Iterating (3.1.15) with $M = 7$ gives

$$X^{(2)}(t) + 0.369X^{(1)}(t) + 0.483X(t) - 1.399 = 0.332W^{(1)}(t)$$

with one-step MSE 0.057 and log-likelihood -0.147

Using these parameters as starting values for maximum likelihood estimation yields the model

$$X^{(2)}(t) + 0.513X^{(1)}(t) + 0.507X(t) - 1.472 = 0.398W^{(1)}(t)$$

where the parameters β, α_0, α_1 and σ have standard errors 0.197, 0.067, 0.118 and 0.034 respectively and the model gives one-step MSE 0.056 and log-likelihood -2.821.

Yeung (1989) and Tong and Yeung (1991) also fit a CAR(2) model to the logged lynx data using maximum likelihood estimation. They give the model

$$X^{(2)}(t) + 0.491X^{(1)}(t) + 0.433X(t) = 0.736W^{(1)}(t)$$

where $X(t)$ denotes the mean-corrected data. It is not clear why their estimates differ from mine, although their calculations may have suffered from the same errors that were apparent in Yeung's estimate of the sunspot model.

Chapter 4

FIRST ORDER CONTINUOUS TIME THRESHOLD AR PROCESSES

The purpose of this chapter is to define a continuous time threshold autoregressive process of order one and discuss some of its properties. The process is defined in Section 4.1 and conditions under which the process is stationary are discussed in Section 4.2. The stationary distribution is derived in Section 4.3. Of particular importance is the behaviour of the process at the thresholds. It will be seen that processes with identical parameters but different boundary conditions have very different properties including different stationary probability densities. Section 4.4 derives the first two moments of the stationary distribution.

Section 4.5 discusses two approximating stochastic differential equations and Section 4.6 describes several approximating Markov chains. One of the approximations given in Section 4.5 will provide the basis for the definition of higher order processes.

4.1 Definition of a CTAR(1) Process

Continuous time threshold autoregressive (CTAR) processes are a generalisation of CAR processes. Instead of the process satisfying a *linear* stochastic differential equation, a CTAR process is defined in terms of a *piecewise linear* stochastic differential equation.

Thus, the first order CTAR process, denoted by CTAR(1), is defined to be a stationary solution of the piecewise linear stochastic differential equation

$$dX(t) + \alpha_0(X(t))X(t)dt + \beta(X(t))dt = \sigma(X(t))dW(t) \tag{4.1.1}$$

where

$$\alpha_0(x) = \alpha_{i,0}, \quad \beta(x) = \beta_i \quad \text{and} \quad \sigma(x) = \sigma_i > 0 \qquad \text{for } r_{i-1} < x < r_i, \tag{4.1.2}$$

and the threshold values $-\infty = r_0 < r_1 < \cdots < r_l = \infty$ partition the real line. If $l = 1$ then $\{X(t)\}$ is a CAR(1) process. The process defined by (4.1.1) and (4.1.2) is a diffusion process with drift coefficient

$$\mu(x) = -\alpha_0(x)x - \beta(x) \tag{4.1.3}$$

and diffusion coefficient

$$\Sigma(x) = \frac{\sigma^2(x)}{2}. \tag{4.1.4}$$

Let $\mathcal{D}(\mathcal{G})$ denote the domain of the generator $\mathcal{G}$ of $\{X(t)\}$. Then, for $f \in \mathcal{D}(\mathcal{G})$, the generator is defined by

$$\mathcal{G}f(x) := \mu(x)f'(x) + \Sigma(x)f''(x), \qquad x \neq r_i \tag{4.1.5}$$

with $\mathcal{G}f(r_i)$ determined by continuity of $\mathcal{G}f(x)$.

Note that the process is not uniquely defined until the behaviour of the process at the thresholds is specified. There is no unique way of doing this. This point is illustrated later in this chapter where different processes with identical parameters but different boundary conditions are considered.

I consider three possible boundary conditions in this chapter, each of which has some useful properties. These are specified by placing conditions on functions in $\mathcal{D}(\mathcal{G})$. For $f \in \mathcal{D}(\mathcal{G})$, the conditions are:

Condition A $\quad \dfrac{\partial f}{\partial x}$ is continuous at $x = r_i$;

Condition B $\quad \sigma_i \dfrac{\partial f}{\partial x}$ is continuous at $x = r_i$;

Condition C $\quad \sigma_i^2 \dfrac{\partial f}{\partial x}$ is continuous at $x = r_i$.

Any one of the boundary conditions A, B or C (and these are not the only possibilities) could be used for modelling purposes. It is important however, in specifying the model as a diffusion process to state which of the boundary conditions is being used. Clearly, if $\sigma(x)$ is continuous there is no distinction between the three boundary conditions. Note that specification of the boundary conditions is critical only when the process is formulated directly as a diffusion process. In Section 4.6, Markov chains which approximate the CTAR(1) process are considered. Each of these automatically implies corresponding boundary conditions for the diffusion limit.

4.2 Conditions for Stationarity

For a stationary solution of a threshold model of order one to exist, one need only consider the sub-models in the outermost regimes. What happens elsewhere does not affect the stationarity of the process. For the application of this principle to SETAR processes, see Tong (1983), and Chan et al. (1985). Each gives conditions involving only the extreme left and extreme right regimes. The principle does not extend to higher order processes: see Tong and Pemberton (1980) and Tong (1989).

For example, a sufficient condition for stationarity is that the sub-models in the outermost regimes are stationary. So a CTAR(1) processes is stationary if the coefficients $\alpha_{1,0}$ and $\alpha_{l,0}$ are both positive.

However, this condition is not necessary. A necessary and sufficient condition for stationarity is that the outermost sub-models force the process towards zero. This leads to the following theorem.

THEOREM 4.2.1 *Suppose that* $\sigma_i > 0,\ i = 1, \ldots, l$. *Then the process* $\{X(t)\}$ *defined by (4.1.1) and (4.1.2) has a stationary distribution if and only if*

$$\lim_{x \to -\infty} \mu(x) > 0 \qquad \textit{and} \qquad \lim_{x \to \infty} \mu(x) < 0.$$

PROOF: If either of the two conditions is violated, if $[a, b]$ is any finite interval and if $0 < \epsilon < (b - a)/2$, then either the expected passage time of $X(t)$ from $b+\epsilon$ to $b-\epsilon$ or the expected passage time from $a-\epsilon$ to $a+\epsilon$ is infinite. On the other hand the expected time, starting from any state $x \in [a+\epsilon, b-\epsilon]$ for $X(t)$

to reach either $a - \epsilon$ or $b + \epsilon$ is bounded. A simple renewal theory argument therefore shows that there can be no stationary distribution for $X(t)$, since such a distribution would necessarily assign probability zero to the interval $[a + \epsilon, b - \epsilon]$ and hence to every finite interval.

□

These conditions are satisfied if and only if one of the following is true:

1. $\alpha_{1,0} > 0$ and $\alpha_{l,0} > 0$;
2. $\alpha_{1,0} > 0$, $\beta_l > 0$ and $\alpha_{l,0} = 0$;
3. $\alpha_{l,0} > 0$, $\beta_1 < 0$ and $\alpha_{1,0} = 0$; or
4. $\alpha_{1,0} = \alpha_{1,0} = 0$, $\beta_1 < 0$ and $\beta_l > 0$.

The last of these represents the situation in which there is Brownian motion at the outermost regimes, each with drift towards zero. In this case, neither of the outermost sub-models is stationary, but the overall threshold model is stationary.

4.3 Stationary Distributions

The distinction between the CTAR(1) processes with different boundary conditions is apparent (unless $\sigma(x)$ is continuous) in the stationary distributions for the different processes.

If both conditions of Theorem 4.2.1 are satisfied I show that there is a stationary distribution by computing it explicitly. To do this, note that $\pi(x)$ is a stationary density for $\{X(t)\}$ if and only if

$$\int_{-\infty}^{\infty} \mathcal{G}f(x)\pi(x)dx = 0 \qquad \text{for all } f \in \mathcal{D}(\mathcal{G}) \tag{4.3.1}$$

(see e.g. Breiman, 1968, p.346). Substituting from (4.1.5) and integrating by parts gives

$$[\mu(x)\pi(x)f(x) + \Sigma(x)\pi(x)f'(x) - \Sigma(x)\pi'(x)f(x)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} [\mu'(x)\pi(x) + \mu(x)\pi'(x) - \Sigma(x)\pi''(x)]\, f(x) = 0$$

for $x \notin \{r_1, \cdots, r_{l-1}\}$ where I have used $\Sigma'(x) = 0$ for $x \neq r_i$ and it is assumed that $\pi(x)$ has continuous second derivatives for $x \neq r_i$. Thus, using the integrability of π, we obtain the forward Kolmogorov equation,

$$\mu(x)\pi'(x) + \mu'(x)\pi(x) - \Sigma(x)\pi''(x) = 0, \qquad x \notin \{r_1, \cdots, r_{l-1}\}, \tag{4.3.2}$$

with boundary conditions,

$$[\mu(x)\pi(x) - \Sigma(x)\pi'(x)]\, f(x) + \Sigma(x)\pi(x)f'(x) \text{ is continuous for all } x. \tag{4.3.3}$$

Integrating (4.3.2) yields

$$\mu(x)\pi(x) - \Sigma(x)\pi'(x) = c_i, \qquad r_{i-1} < x < r_i. \tag{4.3.4}$$

Then integrability of π implies that $c_1 = 0$ and (4.3.3) then implies that $c_i = 0$ for all i. Thus, (4.3.3) reduces to

$$\sigma^2(x)\pi(x)\frac{\partial f}{\partial x} \text{ is continuous for all } x.$$

So under the three boundary conditions given earlier, we obtain corresponding conditions on the stationary density $\pi(x)$:

Condition A $\sigma_i^2\pi(x)$ is continuous at $x = r_i$;

Condition B $\sigma_i\pi(x)$ is continuous at $x = r_i$;

Condition C $\pi(x)$ is continuous at $x = r_i$.

Khazen (1961) considered piecewise linear diffusion processes and used the continuity of the transition probability density as a boundary condition. The above analysis shows that for CTAR(1) processes, this is equivalent to boundary condition C.

Atkinson and Caughey (1968) in their determination of the spectral density for the CTAR(1) process also require continuity of the transition probability density. But since they specify the process with constant variance ($\sigma_i = \sigma$ for all i), there is no distinction between the three boundary conditions considered here.

An expression for the stationary density may be obtained by integrating (4.3.4). This gives

$$\pi(x) = k_i \exp[-\alpha_{i,0}(x + \beta_i/\alpha_{i,0})^2/\sigma_i^2], \quad r_{i-1} < x < r_i, \quad i = 1, \cdots, l, \quad (4.3.5)$$

where $k_1, k_2, \cdots, k_l$ are determined by $\int_{-\infty}^{\infty} \pi(x)dx = 1$ and the boundary conditions above.

Example 4.3.1: Consider the CTAR(1) process

$$dX(t) + 0.50X(t)dt = 0.50dW(t), \qquad \text{if } X(t) < 0,$$
$$dX(t) + X(t)dt = dW(t), \qquad \text{if } X(t) > 0.$$

From (4.3.5) we immediately find that the stationary density of

$\{X(t)\}$ is

$$\pi(x) = \begin{cases} k_1 e^{-2x^2} & x < 0; \\ k_2 e^{-x^2} & x > 0. \end{cases}$$

For boundary condition A, $k_2 = (4\sqrt{2} - 2)/(7\sqrt{\pi})$ and $k_1 = 4k_2$; for boundary condition B, $k_2 = (2\sqrt{2} - 2)/\sqrt{\pi}$ and $k_1 = 2k_2$; and for boundary condition C, $k_1 = k_2 = (4 - 2\sqrt{2})/\sqrt{\pi}$.

Figure 4.3.1 shows the stationary distributions for the process with the different boundary conditions.

Figure 4.3.1: *The stationary density functions for the CTAR(1) process given in Example 4.3.1 with boundary conditions A, B and C.*

Example 4.3.2: Consider the two-threshold CTAR(1) process,

$$\begin{aligned} dX(t) + 0.18X(t)dt &= 1.2dW(t), && \text{if } X(t) < -0.5, \\ dX(t) + 0.50X(t)dt &= dW(t), && \text{if } -0.5 \leq X(t) < 0.5, \\ dX(t) + 0.80X(t)dt &= 0.4dW(t), && \text{if } X(t) \geq 0.5. \end{aligned}$$

Figure 4.3.2: *The stationary density functions for the CTAR(1) process given in Example 4.3.2 with boundary conditions A, B and C.*

From (4.3.5) the stationary density of $\{X(t)\}$ is

$$\pi(x) = \begin{cases} k_1 e^{-0.125x^2} & x < -0.5; \\ k_2 e^{-0.5x^2} & -0.5 < x < 0.5; \\ k_3 e^{-5x^2} & x > 0.5 \end{cases}$$

where $k_2 = ae^{0.09375}k_1$ and $k_3 = be^{1.21875}k_1$. For boundary condition A, $k_1 \approx 0.2039$, $a = 1.44$ and $b = 9$; for boundary condition B, $k_1 \approx 0.2678$, $a = 1.2$ and $b = 3$; and for boundary condition C, $k_1 \approx 0.3107$, and $a = b = 1$.

Figure 4.3.2 shows the stationary distributions for the process with the different boundary conditions.

4.4 Moments of the Stationary Distribution

From (4.3.5), it is possible to derive analytic expressions for the moments of the stationary distribution. The first two moments are given here. These will be useful later in fitting CTAR(1) models to data using the Gaussian likelihood.

The first moment is

$$\begin{aligned}
\mu_1 &= \int_{-\infty}^{\infty} x\pi(x)dx \\
&= \sum_{i=1}^{l} k_i \int_{r_{i-1}}^{r_i} x \exp\left[-\alpha_{i,0}(x+\beta_i/\alpha_{i,0})^2/\sigma_i^2\right] dx \\
&= \sum_{i=1}^{l} k_i \int_{r_{i-1}+\beta_i/\alpha_{i,0}}^{r_i+\beta_i/\alpha_{i,0}} \left(u - \tfrac{\beta_i}{\alpha_{i,0}}\right) \exp\left[\tfrac{-\alpha_{i,0}u^2}{\sigma_i^2}\right] du \\
&= \sum_{i=1}^{l} k_i \left[f\left(r_i + \tfrac{\beta_i}{\alpha_{i,0}}\right) - f\left(r_{i-1} + \tfrac{\beta_i}{\alpha_{i,0}}\right)\right]
\end{aligned}$$

where

$$f(x) = \frac{-\sigma_i^2}{2\alpha_{i,0}} \exp\left[\tfrac{-\alpha_{i,0}x^2}{\sigma_i^2}\right] - \frac{\beta_i\sigma_i\sqrt{\pi}}{\alpha_{i,0}^{3/2}} \Phi\left(x\sqrt{\tfrac{2\alpha_{i,0}}{\sigma_i^2}}\right)$$

and $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

Similarly, the second moment is

$$\begin{aligned}
\mu_2 &= \int_{-\infty}^{\infty} x^2\pi(x)dx \\
&= \sum_{i=1}^{l} k_i \int_{r_{i-1}}^{r_i} x^2 \exp\left[-\alpha_{i,0}(x+\beta_i/\alpha_{i,0})^2/\sigma_i^2\right] dx \\
&= \sum_{i=1}^{l} k_i \int_{r_{i-1}+\beta_i/\alpha_{i,0}}^{r_i+\beta_i/\alpha_{i,0}} (u - \tfrac{\beta_i}{\alpha_{i,0}})^2 \exp\left[\tfrac{-\alpha_{i,0}u^2}{\sigma_i^2}\right] du \\
&= \sum_{i=1}^{l} k_i \left[g\left(r_i + \tfrac{\beta_i}{\alpha_{i,0}}\right) - g\left(r_{i-1} + \tfrac{\beta_i}{\alpha_{i,0}}\right)\right]
\end{aligned}$$

where

$$g(x) = \frac{\sigma_i^2}{\alpha_{i,0}}\left(\frac{\beta_i}{\alpha_{i,0}} - \frac{x}{2}\right)\exp\left(\frac{-\alpha_{i,0}x^2}{\sigma_i^2}\right) + \frac{\sigma_i\sqrt{\pi}}{\alpha_{i,0}^{3/2}}\left(\frac{\sigma_i^2}{2} + \frac{\beta_i^2}{\alpha_{i,0}}\right)\Phi\left(x\sqrt{\frac{2\alpha_{i,0}}{\sigma_i^2}}\right).$$

4.5 Approximating SDEs

4.5.1 CTAR(1) processes with a boundary

Define the CTAR(1) process $\{X_\delta(t)\}$ with boundary width $2\delta > 0$ exactly as the CTAR(1) process defined in (4.1.1) except that the coefficients $\alpha_0(x)$, $\beta(x)$ and $\sigma(x)$ are modified so that they are made continuous in x. Specifically, define

$$\alpha_0(x) = \alpha_{i,0}, \quad \beta(x) = \beta_i, \quad \sigma(x) = \sigma_i > 0 \qquad \text{for } r_{i-1} + \delta < x < r_i - \delta, \tag{4.5.1}$$

and define $\alpha_0(x)$, $\beta(x)$ and $\sigma(x)$ by linear interpolation for $r_i - \delta < x \leq r_i + \delta$.

Thus, $\{X_\delta(t)\}$ is a diffusion process with continuous drift and diffusion coefficients. By Theorem 1 of Gikhman and Skorohod (1969, p.392), the diffusion equation (4.1.1) with coefficients defined by (4.5.1) has a unique solution.

The stationary density $\pi_\delta(x)$ of $\{X_\delta(t)\}$ can be obtained using the general method of Karlin and Taylor (1981, p.221). Integrate (4.3.2) to obtain

$$\sigma^2(x)\pi'_\delta(x) - 2\mu(x)\pi_\delta(x) = k$$

where k is a constant. Multiplying by the integrating factor

$$I(x) = \exp\left(-\int^x \frac{2\mu(u)}{\sigma^2(u)}du\right) = \exp\left(2\int^x \frac{\alpha_0(u)u + \beta(u)}{\sigma^2(u)}du\right)$$

yields

$$\frac{d}{dx}\{\sigma^2(x)I(x)\pi_\delta(x)\} = kI(x).$$

Thus

$$\pi_\delta(x) = \frac{kS(x)}{\sigma^2(x)I(x)} + \frac{c}{\sigma^2(x)I(x)}, \tag{4.5.2}$$

where $S(x) = \int^x I(u)du$. Consideration of the limit of (4.5.2) as $|x| \to \infty$ gives $k = 0$. Thus

$$\pi_\delta(x) = \frac{c}{\sigma^2(x)I(x)} \tag{4.5.3}$$

where c is the uniquely determined constant such that $\int_{-\infty}^{\infty} \pi_\delta(x)dx = 1$.

Now, since $\mu(x)$ and $\sigma^2(x)$ are continuous in x, so is $I(x)$ and thus $\pi_\delta(x)\sigma^2(x)$ is continuous in x. Substituting $\alpha_0(x)$, $\beta(x)$ and $\sigma(x)$ in (4.5.3) we can easily write down $\pi_\delta(x)$ quite explicitly. Here, however, simply observe that the limit of this density as $\delta \to 0$ is given by (4.3.5) where the constants $k_1, k_2, \ldots, k_l$ are determined by $\int_{-\infty}^{\infty} \pi(x)dx = 1$ and boundary condition A.

Note that it is has not been shown that the approximating process converges as $\delta \to 0$ to a CTAR process with boundary condition A but that the limit of the stationary distribution is consistent with its convergence to a process having boundary condition A. Numerical results suggest that the approximating process does converge to the required CTAR process but this has not been shown analytically.

4.5.2 CTAR(1) processes with a white noise approximation

Other approximations lead to different boundary conditions. Let $\{Y_n(t)\}$ be a continuous time Markov process defined on the state space $\{-\sqrt{n}, \sqrt{n}\}$ and with generator $\begin{bmatrix} -n & n \\ n & -n \end{bmatrix}$ and define $I_n(t) = \int_0^t Y_n(u)du$. Then $\{I_n(t)\}$ converges weakly to Brownian motion as $n \to \infty$ (see Brockwell, Resnick and Pacheco-Santiago, 1982). A realisation of these processes is shown in Figure 4.5.1.

Figure 4.5.1: *A possible realisation of* $Y_n(t)$ *and* $I_n(t) = \int_0^t Y_n(u)du$. *As* $n \to \infty$, $I_n(t)$ *converges to Brownian motion. Here,* $n = 100$.

Therefore we may define the approximating process $\{X_n(t), t \geq 0\}$ to be the solution of equation (4.1.1) with $\{W(t)\}$ replaced by $\{I_n(t)\}$. That is

$$dX_n(t) + \alpha_0(X_n(t))X_n(t)dt + \beta(X_n(t))dt = \sigma(X_n(t))Y_n(t).$$

Consideration of the generator of the process $\{(X_n(t), Y_n(t))\}$ as $n \to \infty$ leads

to boundary condition B.

4.6 Discrete Time Approximations

In this section, I consider the problem of approximating the CTAR(1) process $\{X(t)\}$ by a random broken line with vertices defined by a discrete time Markov chain on the times $0 = t_0 < t_1 < t_2 < \cdots$. Such approximations are frequently used in the numerical study of diffusion processes (see, e.g. Gikhman and Skorohod, 1969). Several such approximations are derived; it will be shown that each approximation implicitly determines the boundary conditions.

I consider Markov chain approximations to the stochastic differential equation for a CAR(1) process, generalise them to allow for non-linearity at the thresholds, and obtain the boundary conditions by considering the limiting properties of the generator of the approximating process. The results of Kurtz (1975) guarantee the weak convergence of the approximating process to the CTAR(1) process with the appropriate boundary condition.

One approach is to approximate a CAR(1) process using the Euler method of time discretisation. This yields the Markov chain

$$X_n(t + n^{-1}) = X_n(t) - n^{-1}(\alpha_0 X_n(t) + \beta) + n^{-\frac{1}{2}}\sigma Z_n(t) \tag{4.6.1}$$

where $t = 0, \frac{1}{n}, \frac{2}{n}, \ldots$, $\{Z_n(t)\}$ is a sequence of iid random variables with zero mean and unit variance and $X_n(0)$ is independent of $\{Z_n(t)\}$ and has the same distribution as $X(0)$. Sample paths of $\{X_n(t)\}$ between $t = j/n$ and $t = (j+1)/n$, $(j = 0, 1, 2, \ldots)$ are defined by linear interpolation. Then, with

some conditions on $\{Z_n(t)\}$ (Gikhman and Skorohod, 1969, p.460), the process $\{X_n(t)\}$ will converge weakly to $\{X(t)\}$ on a finite interval of time as $n \to \infty$.

Higher order approximations are available (Platen, 1987) but will not be considered here.

A threshold analogue of (4.6.1) is

$$X_n(t + n^{-1}) = X_n(t) + n^{-1}\mu(X_n(t)) + n^{-1/2}\sigma(X_n(t))Z_n(t), \qquad (4.6.2)$$

where $\mu(x)$ and $\sigma(x)$ are defined in (4.1.2) and (4.1.3). To complete the definition, I shall (arbitrarily) define the values of $\mu(r_i)$ and $\sigma(r_i)$ as $\mu(r_i^+)$ and $\sigma(r_i^+)$ respectively. Here and later I use the notation

$$g(x^+) := \lim_{h\to 0, h>0} g(x+h) \qquad \text{and} \qquad g(x^-) := \lim_{h\to 0, h>0} g(x-h).$$

Two approximations considered below are special cases of (4.6.2) where the distribution of $Z_n(t)$ is specified. It will be shown that these converge to a CTAR(1) process with boundary condition A.

4.6.1 Binary increments

The simplest method is to let $Pr(Z_n(t) = -1) = Pr(Z_n(t) = 1) = \frac{1}{2}$. Then the Gikhman–Skorohod conditions are satisfied. Furthermore, Milshtein (1978) shows that for the CAR(1) approximation given by (4.6.1) the one-step mean-square error

$$E\left\{\left[X_n(t+n^{-1}) - X(t+n^{-1})\right]^2 \mid X(t) = X_n(t) = x\right\}$$

is of order n^{-1} and the one-step mean error

$$E\left\{X_n(t+n^{-1}) - X(t+n^{-1}) \,|\, X(t) = X_n(t) = x\right\}$$

is of order n^{-2}.

Now consider the CTAR(1) approximation given by (4.6.2) where $Z_n(t)$ takes the values ± 1 each with probability $\frac{1}{2}$ and let f be a function in the domain of the generator of $\{X_n(t)\}$. Then

$$\begin{aligned} E\left[f(X_n(s+\delta)) \,|\, X_n(s) = x\right] &= \frac{1}{2} f\left(x + \mu(x)\delta + \sigma(x)\delta\sqrt{n}\right) \\ &\quad + \frac{1}{2} f\left(x + \mu(x)\delta - \sigma(x)\delta\sqrt{n}\right) \end{aligned}$$

where $0 < \delta \leq n^{-1}$ and the two terms on the right hand side correspond to the events $Z_n(t) = 1$ and $Z_n(t) = -1$ respectively. Using the mean value theorem, we obtain

$$\begin{aligned} \frac{E\left[f(X_n(s+\delta)) \,|\, X_n(s) = x\right] - f(x)}{\delta} &= \frac{1}{2}\left(\mu(x) + \sigma(x)\sqrt{n}\right) f'(x + \epsilon_1) \\ &\quad + \frac{1}{2}\left(\mu(x) - \sigma(x)\sqrt{n}\right) f'(x - \epsilon_2), \end{aligned}$$

where, if n is sufficiently large, $0 < \epsilon_1 < \mu(x)\delta + \sigma(x)\delta\sqrt{n}$ and $0 < \epsilon_2 < \mu(x)\delta - \sigma(x)\delta\sqrt{n}$. Hence, the limit of the generator of $\{X_n(t)\}$,

$$\mathcal{G}f(x) = \lim_{n\to\infty} \lim_{\delta\to 0} \frac{E\left[f(X_n(s+\delta) \,|\, X_n(s) = x\right] - f(x)}{\delta},$$

exists if and only if $f'(x)$ is continuous for all x. That is, boundary condition A.

Under this condition, the generator of $\{X_n(t)\}$ is

$$\mu(x)f'(x) + \Sigma(x)f''(x) + O(n^{-\frac{1}{2}})$$

which converges uniformly to the right hand side of equation (4.1.5). Thus, by Theorem 4.29 of Kurtz (1975), $\{X_n(t)\}$ converges weakly to the CTAR(1) process with boundary condition A.

This approximation consists of random broken lines with vertices determined by the Markov chain $\{X_n(j/n);\ j = 0, 1, 2, \ldots\}$. Define for $\frac{j}{n} < t < \frac{j+1}{n}$

$$m(t) = \begin{cases} \mu(X_n(t)) + \sigma(X_n(t))\sqrt{n} & \text{if } Z_n(j/n) = 1. \\ \mu(X_n(t)) - \sigma(X_n(t))\sqrt{n} & \text{if } Z_n(j/n) = -1. \end{cases}$$

Then, the slope of the line segment at time $\frac{j}{n} < t < \frac{j+1}{n}$ is $m(j/n)$. Thus,

$$X_n(\tfrac{j}{n} + \delta) = X_n(\tfrac{j}{n}) + m(\tfrac{j}{n})\delta, \qquad 0 < \delta \leq \tfrac{1}{n}.$$

A natural modification of this approximating process is to introduce an extra vertex whenever the process crosses the threshold and change the slope of the new segment, between the threshold and the next vertex, according to the parameters of the regime entered. Thus, whenever $X_n(t + n^{-1})$ is on the opposite side of the threshold, r_i, from $X_n(t)$, redefine $X_n(t + \delta)$ as

$$X_n^*(t + \delta) = \begin{cases} r_i + m(t + \tau)(\delta - \tau) & 0 < \tau \leq \delta \leq n^{-1} \\ X_n(t) + m(t)\delta & 0 < \delta < \tau \text{ or } 0 = \tau < \delta \end{cases}$$

where $\tau = [r_i - X_n(t)]\,/m(t)$ is the time from t before the threshold is crossed. Using this modification, the same approach as above is followed to obtain the

equation

$$\begin{aligned}\frac{E\left[f(X_n(s+\delta))\,|\,X_n(s)=r_i\right]-f(r_i)}{\delta} &= \frac{1}{2}\left(\mu(r_i^+)+\sigma(r_i^+)\sqrt{n}\right)f'(r_i^++\epsilon_1)\\ &\quad+\frac{1}{2}\left(\mu(r_i^-)-\sigma(r_i^-)\sqrt{n}\right)f'(r_i^--\epsilon_2),\end{aligned}$$

where, if n is sufficiently large, $0<\epsilon_1<\mu(r_i^+)\delta+\sigma(r_i^+)\delta\sqrt{n}$ and $0<\epsilon_2<\mu(r_i^-)\delta-\sigma(r_i^-)\delta\sqrt{n}$.

Thus, the limit of the generator of the modified process exists if and only if

$$\left[\sigma(r_i^+)f'(r_i+\epsilon_1)-\sigma(r_i^-)f'(r_i-\epsilon_2)\right]\to 0 \quad \text{as } n\to\infty.$$

That is, boundary condition B.

A slightly simpler modification to the Markov chain is obtained by multiplying $X_n(t+n^{-1})-r_i$ by $\sigma(X_n(t+n^{-1}))/\sigma(X_n(t))$ whenever $X_n(t+n^{-1})$ as defined by (4.6.2) is on the opposite side of the threshold r_i from $X_n(t)$. This allows for the change in $\sigma(x)$ on crossing the threshold. For example, if $X_n(t)<r_i<X_n(t+n^{-1})$, we redefine $X_n(t+n^{-1})$ as

$$X_n^*(t+n^{-1})=r_i+\left(X_n(t+n^{-1})-r_i\right)\frac{\sigma_{i+1}}{\sigma_i}.$$

Equivalently, let

$$X_n^*(t+\delta)=r_i+m(t)\frac{\sigma_{i+1}}{\sigma_i}(\delta-\tau), \qquad 0<\tau<\delta\le n^{-1}.$$

Then, following the above approach, we also obtain boundary condition B.

4.6.2 Gaussian increments and equal time steps

A common alternative approach is to let $Z_n(t) \stackrel{d}{=} N(0,1)$ in (4.6.1). See, for example, Cox and Miller (1965, p.207) and Ozaki (1985, p.55). Again, the Gikhman–Skorohod conditions are satisfied. For CAR(1) processes, the one-step mean-square error is of order n^{-2} (Milshtein, 1974) and the one-step mean error is of order n^{-2} (Milshtein, 1978).

Now consider the CTAR(1) approximation defined by (4.6.2) where $\{Z_n(t)\}$ denotes an iid sequence of standard normal variates. Then, following the above approach, we obtain for f in the domain of the generator of $X_n(t)$

$$\frac{E\left[f(X_n(s+n^{-1}))\,|\,X_n(s)=x\right]-f(x)}{n^{-1}} = \int_{-\infty}^{\infty}\left[\mu(x)+\sigma(x)z\sqrt{n}\right]f'(x+\epsilon(z))\phi(z)dz$$

where $\epsilon(z) \in \left[0, \mu(x)n^{-1}+\sigma(x)n^{-\frac{1}{2}}z\right]$ and $\phi(z)$ denotes the standard normal density function. Taking the limit as $n \to \infty$, we see that $\mathcal{G}f(x)$ exists for all x if and only if $f'(x)$ is continuous for all x. That is, the approximating sequence gives boundary condition A as $n \to \infty$.

As with the binary increment process, this Markov chain approximation can also be modified to obtain boundary condition B.

4.6.3 Exponential increments and exponential time steps

A more complicated approach is to use random time steps. Let $\{U_k\}$ be a sequence of iid random variables such that $U_k \stackrel{d}{=} \exp(\frac{1}{n})$, let $t_{k+1} = t_k + U_k$ and $Z_n(t_k) = \sqrt{nU_k}$ if k is odd and $Z_n(t_k) = -\sqrt{nU_k}$ if k is even. Then the

CAR(1) process may be approximated by

$$X_n(t_{k+1}) = X_n(t_k) - [\alpha_0 X_n(t) + \beta]\,\Delta t_k + \sigma\sqrt{\Delta t_k} Z_n(t_k) \tag{4.6.3}$$

where $\Delta t_k = U_k$ and $X_n(0)$ is independent of $\{Z_n(t_k)\}$ and has the same distribution as $X(0)$.

The validity of this approximation is demonstrated as follows. Let $\{Y_n(t)\}$ be the continuous time Markov process with state space $\{-\sqrt{n}, \sqrt{n}\}$ defined in Section 4.5.2. Then the length of time $Y_n(t)$ spends in each state is a random variable distributed exponentially with mean $\frac{1}{n}$ and independent of $Y_n(s)$ for all $s \neq t$ (Karlin and Taylor, 1981, pp.145–146). Thus, if $t_1, t_2, \ldots$ denote the times at which Y_t changes state and $t_k < t \leq t_{k+1}$, then

$$I_n(t) = \int_0^t Y_n(u)du = \begin{cases} I_n(t_k) + \sqrt{n}(t - t_k) & \text{if } k \text{ odd;} \\ I_n(t_k) - \sqrt{n}(t - t_k) & \text{if } k \text{ even.} \end{cases}$$

Hence,

$$I_n(t_{k+1}) = I_n(t_k) + \sqrt{n}U_k = I_n(t_k) + \sqrt{\Delta t_k} Z_n(t_k).$$

As discussed in Section 4.5.2, $\{I_n(t)\}$ converges weakly to Brownian motion as $n \to \infty$. So $\sqrt{\Delta t_k} Z_n(t_k)$ can be used to approximate increments of Brownian motion and (4.6.3) converges weakly to a CAR(1) process.

The threshold analogue of (4.6.3) is given by

$$X_n(t_{k+1}) = X_n(t_k) + \mu(X_n(t_k))\Delta t_k + \sigma(X_n(t))\sqrt{\Delta t_k} Z_n(t_k). \tag{4.6.4}$$

Using the same argument as given previously, this process also leads to boundary condition A. The modified process obtained by introducing an extra vertex when the process crosses a threshold leads to boundary condition B.

4.6.4 Random walk

A different approach is available using an approximating random walk. Let X_t be a random walk with steps of size Δx in time intervals of Δt. Let $p(x)$ be the probability of step $+\Delta x$ from state x and $q(x) = 1 - p(x)$ be the probability of step $-\Delta x$ from state x. Then choose step size $\Delta x = k\sigma\sqrt{\Delta t}$ for some $k \geq 1$ and let

$$p(x) = \frac{\Delta t}{2(\Delta x)^2}\left[\sigma^2 - (\alpha_0 x + \beta)\Delta x\right] \tag{4.6.5}$$

$$q(x) = \frac{\Delta t}{2(\Delta x)^2}\left[\sigma^2 + (\alpha_0 x + \beta)\Delta x\right]. \tag{4.6.6}$$

In the simple case that $\Delta x = \sigma\sqrt{\Delta t}$, these simplify to

$$p(x) = \frac{1}{2}\left[1 - \frac{(\alpha_0 x + \beta)\sqrt{\Delta t}}{\sigma}\right]$$

$$q(x) = \frac{1}{2}\left[1 + \frac{(\alpha_0 x + \beta)\sqrt{\Delta t}}{\sigma}\right].$$

The process X_t converges weakly as $\Delta t \to 0$ to a CAR(1) process (see, for example, Cox and Miller, 1965, pp.213–214).

This method may be extended to the CTAR(1) case by defining similar random walks on each regime but with different step sizes. In order to allow for the boundary conditions, it is necessary to include all thresholds in the state space. Thus, step sizes for the outermost regimes can be calculated as in the linear case with $\Delta x_i = \sigma_i\sqrt{\Delta t}$, but in the 'internal' regimes, it is necessary to have a slightly larger step size in order to include the thresholds in the state space.

So for $r_{i-1} < X_t < r_i$, the step size may be chosen as follows:

$$\Delta x_i = \begin{cases} \sigma_i\sqrt{\Delta t} & \text{if } i = 1 \text{ or } i = l; \\ (r_i - r_{i-1}) \Big/ \left\lfloor \frac{r_i - r_{i-1}}{\sigma_i\sqrt{\Delta t}} \right\rfloor & \text{if } 1 < i < l, \end{cases} \tag{4.6.7}$$

where $\lfloor\ \rfloor$ denotes the integer formed by truncation. The transition probabilities for $r_{i-1} < X_t < r_i$ may be obtained from (4.6.5) and (4.6.6). To complete the description of the approximating random walk, we need to define the transition probabilities at the thresholds. Different transition probabilities lead to different boundary conditions.

Consider the case of a CTAR(1) process with one threshold at r. The extension of the following results to allow an arbitrary number of thresholds is quite straight forward.

Approximate the CTAR process $\{X(t)\}$ by a random walk with step sizes $\sigma_i\sqrt{\Delta t}$ and transition probabilities defined by (4.6.5) and (4.6.6). Let p^* be the probability of a step in the positive direction from the threshold r. We first calculate

$$E\left[f(X_{\Delta t}) \,|\, X_0 = r\right] = p^* f(r + \sigma_2\sqrt{\Delta t}) + (1 - p^*) f(r - \sigma_1\sqrt{\Delta t})$$

where f is a function in the domain of $\mathcal{G}$, the generator of $\{X(t)\}$. Thus

$$\begin{aligned} &\frac{E\left[f(X_{\Delta t}) \,|\, X_0 = r\right] - f(r)}{\Delta t} \\ &\qquad = \frac{1}{\Delta t}\left[p^* f(r + \sigma_2\sqrt{\Delta t}) + (1 - p^*) f(r - \sigma_1\sqrt{\Delta t}) - f(r)\right] \\ &\qquad = \frac{\sqrt{\Delta t}}{\Delta t}\left[p^*\sigma_2 f'(r + \epsilon_2) - (1 - p^*)\sigma_1 f'(r - \epsilon_1)\right] \end{aligned}$$

for $0 < \epsilon_1 < \sigma_1 \Delta t$ and $0 < \epsilon_2 < \sigma_2 \Delta t$ by the mean value theorem. Hence,

$$\mathcal{G}f(r) = \lim_{t \downarrow 0} \frac{E\left[f(X_t) \,|\, X_0 = r\right] - f(r)}{t}$$

exists if and only if

$$p^* \sigma_2 f'(r+) = (1 - p^*) \sigma_1 f'(r-). \tag{4.6.8}$$

Now different boundary conditions can be expressed in terms of f by plugging in different values of p^* in (4.6.8). Plugging $p^* = \frac{1}{2}$ into (4.6.8) gives the boundary condition B. Other choices of p^* lead to alternative boundary conditions. For example, $p^* = \frac{\sigma_1}{\sigma_1 + \sigma_2}$ leads to boundary condition A and $p^* = \frac{\sigma_2}{\sigma_1 + \sigma_2}$ leads to boundary condition C.

Chapter 5

HIGHER ORDER CONTINUOUS TIME THRESHOLD AR PROCESSES

This chapter extends some of the results of the previous chapter and defines CTAR processes of order $p > 1$. In Section 5.1, a CTAR(p) process with boundary width $2\delta > 0$ is defined as the first component of a p-dimensional Markov process defined by a stochastic differential equation. The CTAR(1) process with boundary condition A defined in the previous chapter may be considered as the limit of the CTAR(1) process defined here as $\delta \to 0$. Section 5.2 defines a pth order version of the approximating Markov chain with binary increments introduced in Section 4.6.1. Using this approximation equations are derived for conditional expectations. The nature of a regularly spaced sample from a CTAR process is considered in Section 5.3.

5.1 Definition of a CTAR(p) Process

Analysis of the CTAR(p) process when $p > 1$ is more complicated than that of the CTAR(1) process since it involves the analysis of a p-dimensional diffusion process. To overcome the difficulties associated with diffusion processes of order $p > 1$ having discontinuous diffusion and drift coefficients, a CTAR(p) process is defined by generalising the approximation given in Section 4.5.1.

Define the CTAR(p) process $X(t)$ with a boundary width $2\delta > 0$ exactly as the CAR(p) process defined in (2.1.3) except that the parameters $\alpha_0, \alpha_1, \ldots, \alpha_{p-1}$, β and σ are allowed to depend on $X(t)$. Specifically, define $X(t)$ to be the first element of the stationary p-dimensional process $\boldsymbol{S}(t)$ satisfying

$$d\boldsymbol{S}(t) = A(X(t))\boldsymbol{S}(t)dt - \boldsymbol{b}\beta(X(t))dt + \boldsymbol{b}\sigma(X(t))dW(t) \tag{5.1.1}$$

where

$$A(x) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -\alpha_0(x) & -\alpha_1(x) & -\alpha_2(x) & \cdots & -\alpha_{p-1}(x) \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

$$\alpha_j(x) = \alpha_{j,i}, \quad \beta(x) = \beta_i, \quad \sigma(x) = \sigma_i > 0, \qquad \text{for} \quad r_{i-1} + \delta < x \leq r_i - \delta, \tag{5.1.2}$$

and $\alpha_j(x)$, $\beta(x)$ and $\sigma(x)$ are defined by linear interpolation for $r_i - \delta < x \leq r_i + \delta$. The thresholds values $-\infty = r_0 < r_1 < \cdots < r_l = \infty$ partition the real line.

By Theorem 4 of Gikhman and Skorohod (1969, p.402), equation (5.1.1) has a

unique solution $\{\boldsymbol{S}(t)\}$ which is a diffusion process with continuous drift and diffusion coefficients

$$\boldsymbol{\mu}(\boldsymbol{x}) = A(x_1)\boldsymbol{x} - \beta(x_1)\boldsymbol{b} \qquad \text{and} \qquad \Sigma(\boldsymbol{x}) = \frac{\sigma^2(x_1)}{2}\boldsymbol{b}\boldsymbol{b}^T$$

where x_1 denotes the first component of the p-vector, $\boldsymbol{x}$.

As shown in Section 4.5.1, limit as $\delta \to 0$ of the stationary distribution of this process is consistent with its convergence to a process having boundary condition A. The CTAR(1) processes with other boundary conditions have not been generalised to higher orders.

5.2 An Approximating Markov Chain

As with CTAR(1) processes, it will be useful to consider a discrete time Markov chain which approximates the above diffusion process. This approximation will provide a useful tool in analysing and developing inference for CTAR models. I shall use a generalisation of the Euler approximation with binary increments introduced in Section 4.6.1.

Consider the sequence of processes,

$$X_n(t) = [1, 0, \ldots, 0]\boldsymbol{S}_n(t), \qquad t = 0, \frac{1}{n}, \frac{2}{n}, \ldots, \tag{5.2.1}$$

where

$$\boldsymbol{S}_n(t + 1/n) = \boldsymbol{S}_n(t) + n^{-1}\boldsymbol{\mu}(\boldsymbol{S}_n(t)) + n^{-1/2}\sigma(X_n(t))\boldsymbol{b}Z_n(t), \tag{5.2.2}$$

and $\{Z_n(t)\}$ is an i.i.d. sequence such that $Pr(Z_n(t) = -1) = Pr(Z_n(t) = 1) =$

$\frac{1}{2}$. It is also assumed that the initial state $\boldsymbol{S}_n(0)$ has the same distribution as $\boldsymbol{S}(0)$ and is independent of the binary sequence $\{Z_n(t)\}$. The process $\{\boldsymbol{S}_n(t)\}$ defined by (5.2.2) is clearly Markovian.

The argument in Theorem 1 of Gikhman and Skorohod (1969, p.460) may be extended to processes of order $p > 1$ to show that for $\delta > 0$, the finite-dimensional distributions of the process $\{\boldsymbol{S}_n(t), t \geq 0\}$ (with sample paths linear between $t = j/n$ and $t = (j+1)/n$, $j = 0, 1, 2, \ldots$) converge to those of $\{\boldsymbol{S}(t)\}$ as $n \to \infty$.

I have not been able to extend this argument to the case where $\delta = 0$. However, numerical results based on the approximating discrete-time process are found to be essentially the same for all sufficiently small $\delta > 0$ as for $\delta = 0$. Therefore, I shall base all numerical results on the approximating process defined by (5.2.1) and (5.2.2) with $\delta = 0$ without making explicit reference each time to the value of δ. The results can be regarded as applying to a threshold CTAR process with *narrow boundary*.

To use the discrete-time process for modelling, we shall need the conditional expectations,

$$m_n(\boldsymbol{x}, t) := E(X_n(t) \,|\, \boldsymbol{S}_n(0) = \boldsymbol{x}) \qquad \text{and} \qquad m(\boldsymbol{x}, t) := \lim_{n\to\infty} m_n(\boldsymbol{x}, t)$$

and the corresponding higher moments

$$m_{n,j}(\boldsymbol{x}, t) := E(X_n^j(t) \,|\, \boldsymbol{S}_n(0) = \boldsymbol{x}) \qquad \text{and} \qquad m_j(\boldsymbol{x}, t) := \lim_{n\to\infty} m_n(\boldsymbol{x}, t).$$

The mean square prediction errors are then

$$v_n(\boldsymbol{x},t) := m_{n,2}(\boldsymbol{x},t) - m_n^2(\boldsymbol{x},t) \qquad \text{and} \qquad v(\boldsymbol{x},t) := \lim_{n\to\infty} v_n(\boldsymbol{x},t).$$

From (5.2.2), we obtain the backward Kolmogorov equations,

$$\begin{aligned} m_n(\boldsymbol{x},t+n^{-1}) &= \frac{1}{2}m_n\left(\boldsymbol{x}+n^{-1}\mu(\boldsymbol{x})+n^{-1/2}\sigma(x_1)\boldsymbol{b}\ ,\ t\right) \\ &\quad + \frac{1}{2}m_n\left(\boldsymbol{x}+n^{-1}\mu(\boldsymbol{x})-n^{-1/2}\sigma(x_1)\boldsymbol{b}\ ,\ t\right), \quad (5.2.3) \end{aligned}$$

with the initial condition,

$$m_n(\boldsymbol{x},0) = x_1. \quad (5.2.4)$$

The corresponding higher moments, $m_{n,j}(\boldsymbol{x},t)$ satisfy the same equation (5.2.3) and the slightly modified initial condition (5.2.4) with the right hand side replaced by x_1^j.

These equations clearly determine the moments $m_n(\boldsymbol{x},t)$ and $m_{n,j}(\boldsymbol{x},t)$ uniquely.

Alternative approximations of $m(\boldsymbol{x},t)$ and $v(\boldsymbol{x},t)$ may be obtained using the piecewise linear predictors and piecewise constant mean square prediction errors for $X(t_{i+1})$ computed for a linear process depending on the regime in which $X(t_i)$ falls. That is, using (2.1.6) and (2.1.8), we define the approximations

$$m_{PL}(\boldsymbol{x},t) := \boldsymbol{a}^T\left[e^{A(x_1)t}\boldsymbol{x} - \beta(x_1)\int_0^t e^{A(x_1)(t-u)}\boldsymbol{b}du\right], \quad (5.2.5)$$

$$\text{and} \quad v_{PL}(\boldsymbol{x},t) := \boldsymbol{a}^T\left[\sigma(x_1)^2\int_0^t e^{A(x_1)(t-u)}\boldsymbol{b}\boldsymbol{b}^T e^{A^T(x_1)(t-u)}du\right]\boldsymbol{a} \quad (5.2.6)$$

where $\boldsymbol{a} = [1,0,\ldots,0]^T$.

Example 5.2.1: Figure 5.2.1 shows, $m_n([x,0],1)$, the one-step predictors for the CTAR(2) process with boundary $r = 0$ and parameters

$$\alpha_{1,0} = 5.0,\ \alpha_{1,1} = 5.0,\ \sigma_1 = \ 5,\ \beta_1 = 20.0,$$

$$\alpha_{2,0} = 2.0,\ \alpha_{2,1} = 2.0,\ \sigma_2 = 10,\ \beta_2 = 25.0.$$

Notice the deviation from linearity in x of the predictors. Values are given for $n = 10$, $n = 100$ and $n = 500$. The values for $n = 100$ and $n = 500$ are virtually indistinguishable and close also to the corresponding values for $n = 10$.

Figure 5.2.1: *The one-step predictors* $m_n([x,0],1)$ *for the CTAR(2) process defined in Example 5.3.1. Also shown is* $m_{PL}([x,0],1)$.

Figure 5.2.2 shows the one-step mean-squared errors $v_n([x,0],1)$ for the same parameter values. Notice that as we move away from the boundary at $x = 0$, the mean squared errors approach the corresponding values for the processes without any threshold. For n=100

Figure 5.2.2: *The one-step mean square errors* $v_n([x,0],1)$ *for the CTAR(2) process defined in Example 5.3.1. Also shown is* $v_{PL}([x,0],1)$.

and 500 these values are extremely close to the mean-squared errors .399 ($x < 0$) and 6.874 ($x > 0$) for the CAR(2) processes corresponding to the parameters below and above the threshold.

Example 5.2.2: Figure 5.2.3 shows the one-step predictors $m_n(x,1)$ for the CTAR(1) process with boundary $r = 0$ and parameters

$$\alpha_{1,0} = 0.5,\ \beta_1 = 0,\ \sigma_1 = 0.5,$$

$$\alpha_{2,0} = 1.0,\ \beta_2 = 0,\ \sigma_2 = 1.0.$$

Values are plotted for $n = 10$ and $n = 100$ together with a piecewise linear graph showing the one-step predictors for the linear processes defined by the parameter values below and above the threshold. The three graphs are virtually indistinguishable.

Figure 5.2.3: *The one-step predictors $m_n(x,1)$ for the CTAR(1) process given in Example 5.3.2 where $n = 10$ and $n = 100$. Also shown are the one-step predictors, $m_{PL}(x,1)$, for the linear processes defined on each regime.*

Figure 5.2.4: *The one-step mean-squared errors, $v_n(x,1)$ for the CTAR(1) process given in Example 5.3.2 where $n = 10$ and $n = 100$. Also shown are the one-step mean-squared errors, $v_{PL}(x,1)$, for the linear processes defined on each regime.*

Figure 5.2.4 shows the one-step mean-squared errors $v_n(x, 1)$ for the same parameter values. The two constant values plotted above and below the threshold again correspond to the linear processes defined by the parameter values above and below the threshold.

The functions $m(x, t)$ and $v(x, t)$ for $t = 1, 2$ and 5 are shown in Figures 5.2.5 and 5.2.6 respectively. As expected, the values of $m(x, t)$ and $v(x, t)$ for $t > 10$ are very close to the mean, 0, and variance, .35355, of the stationary distribution derived in Example 4.3.1.

Figure 5.2.5: *The best predictors* $m(x, t)$, $-5 \leq x \leq 5$, *for Example 5.3.2 with lead times* $t = 1, 2$ *and 5.*

Notice also that for small t, $m(x, t)$ is reasonably well approximated by the solution $m^*(x, t) = x \exp[-\alpha_0(x)t]$ of the corresponding defining equations with the white noise terms set equal to zero. This solution may also be obtained using the predictors for the linear processes defined on each regime. In general however, this approximation will

Figure 5.2.6: *Conditional mean squared errors* $v(x,t)$, $-5 \leq x \leq 5$, *of the predictors in Figure 5.3.5.*

not be good for large values of the lead time, t.

Figures 5.2.7 and 5.2.8 illustrate the difference between the CTAR(1) process with the same parameters but different boundary conditions. The difference between the conditional means (Figure 5.2.7) is relatively small. However the difference between the one-step mean-squared errors is more pronounced, as seen in Figure 5.2.8.

5.3 Regularly Sampled CTAR Processes

As discussed in Section 2.5, a sample from a linear CAR(p) process taken at regularly spaced time intervals follows a discrete time ARMA($p, p-1$) process. This leads one to consider whether an equally spaced sample from a CTAR(p)

Figure 5.2.7: *The one-step predictors for the CTAR(1) process given in Example 5.3.2 with different boundary conditions.*

Figure 5.2.8: *The one-step mean squared errors for the CTAR(1) process given in Example 5.3.2 with different boundary conditions.*

process also has a simple discrete time interpretation. Unfortunately, this is not the case.

Of course, when a CTAR process remains within a regime for a long time, its behaviour will resemble that of a CAR(p) process, and equally spaced observations taken during that time will therefore appear to follow an ARMA($p, p-1$) model. If it is assumed that a CTAR process can cross a regime no more than once between observations, then the sampled process will be 'close' to a discrete time threshold process where the sub-model in each regime is an ARMA(p,$p-1$) model. However, the result is not exact since discrete time models allow model changes only at the times of observation whereas a CTAR process may cross a threshold between observations. Furthermore, assuming the process crosses a regime no more than once between observations is unrealistic as will be shown in the next chapter.

Chapter 6

CONTINUOUS TIME THRESHOLD AUTOREGRESSIVE MODELLING

In this chapter, I consider the problem of estimating a CTAR model from observations $\{x(t_i)\}$ at the discrete times $t_1, \ldots, t_N$. Given observations of a CTAR process where the order and number of regimes is specified, we are concerned with the computation of one-step forecasts, the corresponding mean squared errors and the related problem of parameter estimation by maximisation of the Gaussian likelihood. This is also the approach of Yeung (1989) and Tong and Yeung (1991) who make the approximation that between successive observations in the same regime, the process evolves as the linear process appropriate to that regime. Tong and Yeung's approach is described in Section 6.1. Sections 6.2 and 6.3 develop a different approach based on the discrete time Markov chain approximation described in Chapter 5. A number of simulated and real data sets are analysed and the results compared with those obtained from the method of Tong and Yeung.

The modelling paradigm for CAR processes outlined in Section 3.3 is extended in Section 6.4 to CTAR processes. This involves selection of order, testing

for non-linearity, selection of the number of regimes, model estimation and diagnostic checking. Forecasting is discussed separately in Section 6.5. It is suggested that highest density regions should be used as forecast regions.

6.1 The Tong–Yeung Approach to Estimation

Tong and Yeung (1991) developed an estimation procedure for CTAR models by modifying the maximum likelihood estimation algorithm for CAR models described in Section 3.2. Their approach was first mentioned in Tong (1989) and is also discussed in detail by Yeung (1989). Tong (1990, p.321) is a further reference.

In modifying the algorithm for estimating CAR processes, they make two approximations:

1 if $x(t_i)$ and $x(t_{i+1})$ lie in the same regime, then $x(t)$ remains within that regime during the period (t_i, t_{i+1}); and

2 if $x(t_i)$ and $x(t_{i+1})$ lie in different regimes, then the integral path of $x(t)$ crosses the threshold between the regimes once and once only during the time interval (t_i, t_{i+1}).

The Tong–Yeung method is analogous to that of Section 3.2 except that it uses the piecewise linear predictors and piecewise constant mean square prediction errors for $X(t_{i+1})$ computed from one of l linear processes depending on which regime $x(t_i)$ falls. That is, they apply the Kalman recursions with model parameters changing according to the value of $x(t_i)$ in order to compute the one-step predictor of $X(t_{i+1})$ and its mean square error.

They also make an adjustment when $x(t_i)$ and $x(t_{i+1})$ are on opposite sides of a threshold. I have omitted this adjustment when implementing their method, since by allowing the predictor of $X(t_{i+1})$ and its mean squared error to depend on $x(t_{i+1})$, artificially low one-step mean squared errors were obtained and the estimators were not as good as those obtained without the adjustment.

Simulations show that the two assumptions on which Tong and Yeung base their method are unrealistic. Figure 6.1.1 gives two examples of simulated CTAR(1) processes. The processes were simulated using the Markov chain approximation

$$X_n(t + n^{-1}) = X_n(t) + n^{-1}\mu(X_n(t)) + n^{-1/2}\sigma(X_n(t))Z_n(t) \qquad (6.1.1)$$

where $Pr(Z_n(t) = -1) = Pr(Z_n(t) = 1) = \frac{1}{2}$, $n = 1000$ and $\delta = 0$ (see Section 4.6.1 for details). The parameters used were those defined in Examples 4.3.1 and 4.3.2. The process was observed at unit intervals; the observation points are circled. Clearly, Tong and Yeung's assumptions are violated for these processes. Between most observations, the process switches regimes many times. This occurs for consecutive observations that lie in the same regime and for consecutive observations that lie in different regimes. In the second example, there are cases where the observations lie in adjacent regimes and the process travels into the third regime between the observations. There are also cases where the observations do not even lie in adjacent regimes. These examples are not pathological but typical of many CTAR(1) processes I have simulated. Higher order processes are often smoother than first order processes, but even then simulations indicate that the above assumptions are dangerous.

Tong and Yeung are, of course, aware of the problems inherent in their assump-

Figure 6.1.1: *Simulated CTAR(1) processes demonstrating that the processes may cross the thresholds many times between observations. Observations are marked with circles at unit intervals.*

tions. Tong and Yeung (1989) say "whether the above assumption is valid or not clearly depends on the data at hand" and later "the application of continuous time TAR models is somewhat limited due to the unknown switching conditions." Tong (1990, p.321) goes further and states that "it is intuitively clear that the assumptions above will be satisfied if the sampling intervals are sufficiently small." However, Figure 6.1.1 shows that even when the sampling intervals are very small (say 0.1), the assumptions may be violated. In practice the observer may not be able to choose the sampling interval.

The next two sections introduce estimation methods which do not place assumptions on the switching conditions between observations. The resulting estimates will be compared to those obtained using Tong and Yeung's method.

6.2 Estimation of CTAR(1) Models

Define the conditional expectations

$$\begin{aligned} m(x,t) &= E\left[X(t) \,|\, X(0) = x\right], \\ m_2(x,t) &= E\left[X^2(t) \,|\, X(0) = x\right], \\ v(x,t) &= m_2(x,t) - m^2(x,t), \\ m_i &= m\left(x(t_{i-1}), t_i - t_{i-1}\right) \\ \text{and} \quad v_i &= v\left(x(t_{i-1}), t_i - t_{i-1}\right) \end{aligned}$$

where $i \geq 2$. Thus, m_{i+1} denotes the one-step predictor of $X(t_i)$ and v_{i+1} denotes the one-step mean squared error. Our aim will be to maximise the

Gaussian likelihood of the data, namely

$$L(\boldsymbol{\theta}; x(t_1), \ldots, x(t_N)) := (2\pi)^{-N/2}(v_1 v_2 \cdots v_N)^{-1/2} \exp\left\{-\sum_{i=1}^{N} \frac{(x(t_i) - m_i)^2}{2v_i}\right\},$$

where $\boldsymbol{\theta}$ denotes the vector of parameters to be estimated, or equivalently to minimise

$$-2 \log L = N \log(2\pi) + \sum_{i=1}^{N} \log v_i + \sum_{i=1}^{N} \frac{(x(t_i) - m_i)^2}{v_i}. \tag{6.2.1}$$

We define $m_1 := EX(t_1)$ and $v_1 := \mathrm{Var}(X(t_1))$. These may be calculated for given parameter values $\alpha_{i,0}$, β_i and σ_i $(i = 1, \ldots, l)$ and threshold values r_i, $(i = 1, \ldots, l-1)$ using the results of Section 4.4.

Then m_i and v_i $(i \geq 2)$ may be calculated using the piecewise linear approximations

$$m_i \approx m_{PL}\left(x(t_{i-1}), t_i - t_{i-1}\right) \quad \text{and} \quad v_i \approx v_{PL}\left(x(t_{i-1}), t_i - t_{i-1}\right)$$

given by equations (5.2.5) and (5.2.6). I shall refer to the estimation using these approximations as Method A.

An alternative approach to calculating m_i and v_i is to use the approximations

$$m_i \approx m_n\left(X(t_{i-1}), t_i - t_{i-1}\right) \quad \text{and} \quad v_i \approx v_n\left(X(t_{i-1}), t_i - t_{i-1}\right)$$

where $n = k/(t_i - t_{i-1})$ for a large integer k. These are obtained by solving (5.2.3) and the analogous equation for the second moment, $m_{n,2}(x, t)$. I shall refer to the estimation using these approximations as Method B.

The parameters $\alpha_{i,0}$, β_i and σ_i and threshold values r_i are then estimated by

minimising (6.2.1) using a non-linear optimisation algorithm.

For the CTAR(1) process defined in Example 4.3.1, these approximations are illustrated in Figures 5.2.3 and 5.2.4. Clearly, Method B provides a more accurate approximation.

Also, since the piecewise linear approximation is the same for CTAR(1) processes with different boundary conditions, it is clear that Method A cannot distinguish between these processes while Method B can be used (by appropriate adjustment of the difference equations) to fit models with different boundary conditions. In fact, it is difficult to define the model that is fitted using Method A since, when two observations lie in different regimes, the model parameters don't change until the time of observations rather than when the process crosses the threshold. In this sense, the models fitted using Method A to equally spaced data behave like a continuous time version of a discrete time threshold model with delay equal to the observation spacing. To distinguish the models estimated using Method A from CTAR models, I shall call them 'piecewise linear CAR' models.

However, Method A is much faster computationally. Because of its speed, Method A may be a useful source of preliminary estimates with which to initialise subsequent likelihood maximisation with Method B. The calculation of the Gaussian likelihood by Method B can of course be improved by increasing the chosen value of k, but only at the expense of increased computational time.

Tong and Yeung's method is similar to Method A. The differences are:

1 they make an adjustment for threshold crossing (see previous section);

2 they drop the first term (involving m_1 and v_1) from the sums (and so calculate the likelihood conditional on the first observation);

3 they do not include drift parameters, β_i;

4 they only consider models with two regimes ($l = 2$); and

5 they restrict r to the set $\{\bar{x} \pm qs\}$ where $q = 0, \frac{1}{7}, \ldots, 1$.

Example 6.2.1: Twenty independent realisations of the CTAR(1) process defined in Example 4.3.1 were generated using the Markov chain (6.1.1) with $n = 1000$ and $\delta = 0$. Each realisation consisted of 100 observations taken at unit time intervals apart. For each realisation, the parameters were estimated using Methods A and B; for Method B, $k = 10$ was used. The results are given below.

Estd. parameter	r	$\alpha_{1,0}$	β_1	σ_1	$\alpha_{2,0}$	β_2	σ_2
Actual value	0.000	0.500	0.000	0.500	1.000	0.000	1.000
Sample mean A	0.016	0.706	0.099	0.580	2.796	−0.245	1.113
Sample st.dev. A	0.046	0.408	0.122	0.099	4.998	0.917	0.668
Sample mean B	−0.044	0.503	0.004	0.467	1.147	−0.074	0.921
Sample st.dev. B	0.151	0.174	0.143	0.158	0.307	0.223	0.146

Both methods give reasonably good estimates of the true parameters. However, the variability of the estimated parameters was often substantially larger for Method A and Method A was notably less successful in estimating the AR coefficients. Method A did however locate the threshold well in each case, possibly because of the sharp discontinuity in the one-step prediction mean squared error at the boundary in the piecewise linear approximation, which makes the corresponding Gaussian likelihood sensitive to changes in r.

Maximising the likelihood using Method A was numerically difficult. Different starting values for the parameters often led to different estimates. Also, in a number of cases the parameters diverged. These cases have been dropped in the above summary statistics. These difficulties are probably due to the discontinuity noted above. No such difficulties were encountered using Method B which has a smooth mean squared error function.

Example 6.2.2: I next consider the modelling of the IBM closing stock prices discussed in Section 3.4.1. Again, the relative daily price changes are modelled and the data are treated as uniformly spaced.

A model of order one with two regimes is estimated. The estimated model parameters are shown below for the models fitted using Methods A and B, together with the values of -2 times the corresponding log likelihood and the observed one-step mean squared errors. The linear model is the maximum likelihood CAR(1) model for the data found in Section 3.4.1. As before, $k = 10$ was used for Method B.

Estd. parameter	r	β_1	$\alpha_{1,0}$	σ_1	β_2	$\alpha_{2,0}$	σ_2
Linear model		–0.11	1.56	1.69			
Piecewise Model A	–1.30	–1.44	5.11	5.49	–0.01	1.15	1.37
CTAR(1) Model B	–0.56	0.84	2.87	2.80	0.61	0.43	1.15

	$-2\log(L)$	MSE
Linear model	590.6	0.879
Piecewise Model A	566.8	0.847
CTAR(1) Model B	578.3	0.840

Although the non-linear models show a substantial increase in the Gaussian likelihood (as compared with the linear model), the reduction in mean squared error (the average of the squares of the observed one-step prediction errors) is quite small. There are still gains to be made however from the non-linear models. Unlike linear models, they have the property that forecasting mean squared errors *depend on the level of the process* (cf. Pemberton, 1989 and Tong, 1990, p.349). Figure 6.2.1 shows the one-step predictors and their mean squared errors of both models as functions of x. It clearly indicates those values of x for which the model one-step mean squared error is small. Simulations with Model B are found to give marginal distributions quite similar to the empirical marginal distribution of the data.

Figure 6.2.1 shows that the piecewise model mean squared error function is quite different from that of the CTAR(1) model. Even a piecewise constant approximation to the mean squared error of the CTAR(1) is not good over the range of observed values (it has the values 1.36 for $x < -.56$ and 0.89 for $x > -.56$). The piecewise constant approximation is also poor for the corresponding CTAR(1) process with the parameters of Model A.

The Gaussian likelihood and mean squared error for Model A are computed with the piecewise linear approximation described above. These quantities are exact for the discrete-time Markov process with the appropriate Gaussian initial distribution and Gaussian one-step transition probabilities determined (for the time interval $[t, t+1]$) using one of two CAR(1) processes, chosen in accordance with the

Figure 6.2.1: *One-step predictors and mean squared errors for model A of the IBM closing stock data*

value of $X(t)$. Although this process is a very good model for the data (as indicated by the corresponding Gaussian likelihood and mean squared error), it differs substantially from a CTAR(1) process with narrow boundary and the same parameters, because of the deviation from piecewise constancy of the mean squared error for the latter.

As indicated previously we could, at the expense of computer time, improve on the calculation of the Gaussian likelihood in the fitting of Model B by choosing a value of k greater than 10.

Both models identify a threshold with negative value, confirming the conclusion of Tong and Yeung regarding this series. As they point out, this suggests that the dynamics of the closing stock prices change according as the stocks are falling at greater or less than a critical relative rate.

Example 6.2.3: The data for this example are temperatures (in degrees Celsius) from the top of a BHP steel blast furnace. The temperatures were recorded every six minutes for a period of almost ten days giving 2389 observations; ten observations are missing. The data set is given in the appendix. A graph of the series (Figure 6.2.2) shows the temperatures usually lie in the range 70–200 ^oC but there are occasional bursts of very high temperatures up to 400 ^oC. This pattern appears to be well-suited to threshold-type modelling. From the graph, a threshold around 190 ^oC would seem to be appropriate.

A model of order one with two regimes was estimated. A time scale of one unit for each six minutes was used. The estimated model parameters are shown below for the models fitted using Methods A

and B, together with the values of -2 times the corresponding log likelihood and the observed one-step mean squared errors. The linear model is the maximum likelihood CAR(1) model for the data. As before, $k = 10$ was used for Method B.

Estd. parameter	r	β_1	$\alpha_{1,0}$	σ_1	β_2	$\alpha_{2,0}$	σ_2
Linear model		–17.96	0.1270	16.23			
Piecewise Model A	195.4	–21.98	0.1569	14.22	–22.04	0.1284	53.15
CTAR(1) Model B	176.6	–10.18	0.0642	13.02	20.51	0.0577	47.48

	$-2\log(L)$	MSE
Linear model	19798.4	232.1
Piecewise Model A	19253.8	231.3
CTAR(1) Model B	19242.8	236.7

Figure 6.2.2: *Plot of temperatures from top of blast furnace. The threshold value from Model B is also shown.*

Simulated observations from Model B were generated using the Markov chain (6.1.1) with $n = 100$ and $\delta = 0$. One thousand obser-

vations were taken from the simulated series at unit time intervals apart. The observations, plotted in Figure 6.2.3, show the model has successfully captured the essential elements of the system dynamics.

Figure 6.2.3: *Plot of simulated series using Model B.*

6.3 Estimation of CTAR(p) Models

When $p = 1$ it is a straightforward matter to compute the Gaussian likelihood of the observations $x(t_1), \ldots, x(t_N)$ to a high degree of accuracy by choosing a suitably large value of the parameter n of the approximating process defined by (6.1.1). When $p > 1$ the problem is more difficult since the process $\{X(t)\}$ is no longer Markov.

One approach to this difficulty is to use a piecewise linear approximation, applying the linear Kalman recursions with model parameters changing accord-

ing to the current value of $x(t_i)$ in order to compute the one-step predictor of $X(t_{i+1})$ and its mean squared error. This is essentially the method of Tong and Yeung. It is the extension of Method A of the previous section to higher dimensions. Although this is a convenient and fast algorithm, the assumptions on which it is based are unrealistic (see Section 6.1) and it is difficult to give a precise interpretation to the Gaussian likelihood computed in this way when $p > 1$.

The approach followed here is an extension of Method B. I use a set of recursions which determine the *exact* likelihood of the observations, $x(t_1), \ldots, x(t_N)$, in terms of the distribution of $\boldsymbol{S}(t_1)$ and the transition density of the Markov process $\{\boldsymbol{S}(t)\}$. The Gaussian likelihood is computed by replacing the transition density by a Gaussian density with the same moments of orders 1 and 2. These moments are computed numerically with the aid of the approximating process defined by (5.2.1) and (5.2.2).

In order to obtain the required recursions, it is convenient to introduce a slightly different notation, writing the state vector $\boldsymbol{S}(t)$ as

$$\boldsymbol{S}(t) = \begin{bmatrix} X(t) \\ \\ \boldsymbol{U}(t) \end{bmatrix},$$

where $\boldsymbol{U}(t)$ is the $(p-1) \times 1$ vector consisting of the first $p-1$ derivatives of $X(t)$.

Consider now the joint probability density, f_r, of $X(t_r), \boldsymbol{U}(t_r), X(t_{r-1})$, $X(t_{r-2}), \ldots, X(t_1)$. From the Markov property of $\{\boldsymbol{S}(t)\}$ it is easy to check

that

$$f_{r+1}(x_{r+1}, \boldsymbol{u}_{r+1}, x_r, x_{r-1}, \ldots, x_1) = \int p(x_{r+1}, \boldsymbol{u}_{r+1}, t_{r+1} - t_r \,|\, x_r, \boldsymbol{u}_r) f_r(x_r, \boldsymbol{u}_r, x_{r-1}, \ldots, x_1) d\boldsymbol{u}_r, \quad (6.3.1)$$

where $p(x_{r+1}, \boldsymbol{u}_{r+1}, t_{r+1} - t_r \,|\, x_r, \boldsymbol{u}_r)$ denotes the pdf of $(X(t_{r+1}), \boldsymbol{U}^T(t_{r+1}))^T$ given $\boldsymbol{S}(t_r) = (x_r, \boldsymbol{u}_r^T)^T$. For a given set of observed values $x_1, \ldots, x_N$ at times $t_1, \ldots, t_N$, the functions $f_2, \ldots, f_N$ are functions of $\boldsymbol{u}_2, \ldots, \boldsymbol{u}_N$ respectively. These functions can easily be computed recursively from (6.3.1) in terms of f_1 and the functions $p(x_{r+1}, \cdot, t_{r+1} - t_r \,|\, x_r, \cdot)$. The likelihood of the observations $x_1, \ldots, x_N$ is then

$$L(\boldsymbol{\theta}; x_1, \ldots, x_N) = \int_{\boldsymbol{u}_N} f_N(\boldsymbol{u}_N) d\boldsymbol{u}_N. \quad (6.3.2)$$

In all the numerical calculations given below I have taken $f_1(x, \boldsymbol{u})$ to be the Dirac delta function assigning mass one to $(x_1, \boldsymbol{0}^T)^T$. This means that the likelihood in (6.3.2) is actually that of $x_2, \ldots, x_N$ *conditional* on $X(t_1) = x_1$ and $\boldsymbol{U}(t_1) = \boldsymbol{0}$. The first and second order moments of the transition density $p(x_{r+1}, \boldsymbol{u}_{r+1}, t_{r+1} - t_r \,|\, x_r, \boldsymbol{u}_r)$ are found using the approximating process defined by (5.2.1) and (5.2.2) with $n = 10/(t_{r+1} - t_r)$. The 'Gaussian likelihood' is then found by replacing the transition densities by Gaussian densities with the same first and second order moments. The integrals in the recursions (6.3.1) are replaced by approximating Riemann sums.

Notice that the filtered value of the unobserved vector $\boldsymbol{U}(t_r)$, $r = 1, \ldots, N$, (i.e. the conditional expectation of $\boldsymbol{U}(t_r)$ given $X(t_i) = x_i$, $i = 1, \ldots, r$) is

readily obtained from the function f_r as

$$\tilde{\boldsymbol{u}}_r = \frac{\int \boldsymbol{u} f_r(\boldsymbol{u}) d\boldsymbol{u}}{\int f_r(\boldsymbol{u}) d\boldsymbol{u}}. \tag{6.3.3}$$

On the other hand, the calculation of the expected value of $X(t_{r+1})$ given $X(t_i) = x_i$, $i = 1, \ldots, r$ involves a much more complicated higher dimensional multiple integration. An alternative natural predictor of $X(t_{r+1})$ which is easy to compute can be found from

$$x_{r+1} = m\left([x_r, \tilde{\boldsymbol{u}}_r^T]^T,\ t_{r+1} - t_r\right), \tag{6.3.4}$$

where $m\left([x_r, \tilde{\boldsymbol{u}}_r^T]^T,\ t_{r+1} - t_r\right)$ is approximated by $m_n\left([x_r, \tilde{\boldsymbol{u}}_r^T]^T,\ t_{r+1} - t_r\right)$ as defined by equation (5.2.3).

There are two natural criteria for model-fitting based on the calculations just described:

B1: maximisation of the Gaussian likelihood and

B2: minimisation of the mean squared error of the predictors defined by (6.3.4).

The resulting models will be referred to as Models B1 and B2 respectively. Model A is that obtained by the piecewise linear method. An important feature of both Models B1 and B2 is that the order of approximation can be increased arbitrarily (at the expense of increasing computation time).

Example 6.3.1: The above methods of estimation are illustrated via Wolf's sunspot numbers considered in Section 3.4.2. Here the

data used are the sunspot numbers for the years 1770–1869 as given in Brockwell and Davis (1991, p.6). The parameter estimates are shown in the following table. In each case, the computed likelihoods are conditional on $X(t_1) = x(t_1) = 101$ and $U(t_1) = 0$. The optimisation was constrained for all models so that the roots of $z^2 + \alpha_{1,i}z + \alpha_{0,i} = 0$, $i = 1, 2$ all have negative real parts.

Estd. parameter	r	β_1	$\alpha_{0,1}$	$\alpha_{1,1}$	σ_1	β_2	$\alpha_{0,2}$	$\alpha_{1,2}$	σ_2
Linear model		-21.0	0.433	0.495	24.7				
Piecewise model A	48.8	-23.0	0.68	0.54	23.5	-11.1	0.19	0.00	12.7
CTAR(2) Model B1	10.0	-31.6	0.33	8.74	43.3	-23.0	0.46	0.55	28.4
CTAR(2) Model B2	36.3	-17.3	0.34	0.21	7.10	-21.8	0.44	0.70	42.3

	$-2\ln(L)$	MSE
Linear model	823.6	239.1
Piecewise model A	792.2	176.5
CTAR(2) Model B1	796.6	215.2
CTAR(2) Model B2	939.6	177.5

As a check on the fitted models, simulations of each of the CTAR(2) processes with the parameters of models A, B1 and B2 were calculated. Each simulation consisted of 2000 observations and the empirical marginal distributions of the simulated observations were compared with the marginal distribution of the sunspot numbers themselves. The results are shown in Figure 6.3.1. Clearly, there is a strong resemblance between the marginal distribution for model B1 and the empirical distribution of the sunspot numbers.

As noted previously, one-step predictors and mean squared errors for model A are computed using a fast, well-defined algorithm. However,

Figure 6.3.1: *Histograms of the sunspot numbers and simulated observations from the CTAR(2) models A, B1 and B2.*

the resulting likelihood does not have the same clearly-defined interpretation it has when $p = 1$. In fact the CTAR *model* with these parameters is a poor fit to the data (in terms of Gaussian likelihood, one-step mean squared error and marginal distribution). There is a generalised non-linear regression model implicit in Model A however in which $EX(t_n)$ is a complicated n-dependent function of all the preceding observations. Although this model has no clear relation to a non-linear continuous time autoregression, it is a very good model for the sunspot numbers as can be seen from the tables above.

The forecasting performance of the four models were compared by using them to compute one-step forecasts of the annual sunspot numbers for the years 1870–1889 (these can be found in Tong, 1990, p.470). The observed mean squared errors for the models were

	Observed MSE
Linear model	469.8
Piecewise model A	391.5
CTAR(2) Model B1	450.9
CTAR(2) Model B2	419.3

Of the non-linear models, Model A gives the greatest reduction in mean squared error. Although the relation between Model A and continuous-time autoregression is only tenuous, its success here suggests that it is worthy of further study as a practical tool for non-linear prediction.

6.4 Modelling Paradigm

6.4.1 Tests for non-linearity

Tong and Yeung (1990) have developed tests to detect threshold non-linearity in continuous time autoregressive processes. Their approach follows that which is used for detecting discrete time threshold non-linearity using arranged autoregression.

A linear model is fitted using the Kalman recursions and the innovations $I(t_i)$ are stored. These are then ordered according to the magnitude of the observations $X(t_i)$. Thus, if there is threshold non-linearity, the innovations are grouped according to the regime in which they fall. Tong and Yeung consider tests based on CUSUMS and the supposed orthogonality of the innovations to detect non-linearity.

6.4.2 Fitting a model

Given that the data shows evidence of threshold non-linearity, the results of the previous sections and chapters may be used to find a threshold model.

The approach adopted here is to first fit a linear model following the modelling paradigm of Chapter 3. Having selected the appropriate order and fitted a linear CAR model using the Yule–Walker estimates, a threshold can be added and the likelihood of the new model maximised. Thresholds are added until the AIC increases. Each new thresholds is added using the mid-point of the largest regime of the current model as a starting value.

This modelling approach led to the CTAR(1) model with one threshold for the IBM data given in Example 6.2.2.

6.4.3 Diagnostics

One-step predictors may be found using the function $m_n(\boldsymbol{x}(t_{i-1}), t_i - t_{i-1})$ and so innovations may be calculated as $I(t_i) = X(t_i) - m_n(\boldsymbol{x}(t_{i-1}), t_i - t_{i-1})$. For models of order $p > 1$, the state vector $\boldsymbol{x}(t_{i-1})$ may be computed using numerical derivatives such as central differences.

These innovations should be orthogonal to past observations but not necessarily Gaussian. Therefore, tests for serial correlation can be applied as for discrete time models (see Brockwell and Davis, 1991, Section 9.4).

6.5 Forecasting

The functions $m_n(\boldsymbol{x}, t_N + h)$ and $v_n(\boldsymbol{x}, t_N + h)$ may be used to obtain the mean and variance of the forecasts. However, consideration needs to be given to obtaining appropriate forecast estimates and prediction regions rather than simply applying linear results to the rich environment of non-linear models. Simply reporting the mean and variance of the forecast is of limited value when the forecast density is non-Gaussian. For example, the forecast distribution of a CTAR process may be asymmetric or multi-modal. In this case, the prediction interval calculated using the mean plus and minus two standard deviations is inappropriate and even misleading. The probability of the actual process falling within the interval may be much less than the supposed 95%.

Yet this approach is commonly used for discrete time threshold modelling (e.g. Tong and Moeanaddin, 1988).

Further information about the forecast distribution may be obtained by generating many independent realisations of the process using the approximating sequence defined by (5.2.1) and (5.2.2). Suppose we wish to forecast the series at time $T+h$ based on observations taken at times up to T. The distribution of $X(T+h)$ given observations up to time T may be studied by simulating the process between times T and $T+h$. The initial state for each realisation may be constructed using the observation at time T and numerical derivatives of the process at time T obtained using forward differences. Let Y_i denote the value of the ith realisation of $X(T+h)$ calculated in this way. Then $\{Y_i\}$ may be used to obtain appropriate point and interval estimates.

I suggest a good forecast region must contain points where the density is relatively high. That is, the highest density region (HDR) which is defined below.

DEFINITION 6.1 *A* $100(1-\alpha)\%$ *HDR is equivalently defined as*

1. *the* $100(1-\alpha)\%$ *confidence region in which every point has probability density at least as large as that of any point outside the region; or*
2. *the* $100(1-\alpha)\%$ *confidence region occupying the smallest possible volume in the sample space.*

In the case of a multi-modal distribution, a HDR often consists of several disjoint sub-intervals.

A common objection to HDRs is the lack of suitable and efficient software. However, Chapter 7 provides an algorithm which enables the rapid calculation of a HDR given any univariate continuous density.

A highest density prediction region may be constructed from the empirical density of $\{Y_i\}$.

An important difference between non-linear and linear prediction regions is monotonicity. It is well known that for linear models, the further ahead we forecast, the less 'reliable' is the forecast. However, this result is not generally true for non-linear models; the forecast variance does not always increase in time. For example, the process may be cyclic with greater variability at the tops of peaks than at the bottoms. Then the forecast variance may be smaller for forecasts near peaks than for forecasts near troughs. Thus, it is possible to construct an example in which a long-term forecast from a given point is more reliable than a short-term forecast from the same point. This phenomenon is documented by Tong and Moeanaddin (1988), Pemberton (1989) and Tong (1990, p.349).

Example 6.5.1: This approach to forecasting was applied to the sunspot numbers considered in Example 6.3.1. Using model B1, the sunspot numbers for the years 1870–1879 were forecast by generating 1000 independent realisations of the process using the approximating sequence defined by (5.2.1) and (5.2.2). The initial state for each realisation was constructed using the observation from 1869 and the numerical derivative obtained using a forward difference. Then the empirical density of the simulations was estimated (using the Gaussian kernel approach) and the 95% HDR was computed. The sample

mean of the simulations was used as a point estimate.

Figure 6.5.1: *Forecasts of the sunspot numbers showing the 50% HDR and 95% HDR of the forecast density. The solid line represents the observed values and the dashed line the mean of the forecast density.*

Figure 6.5.1 shows the forecast mean and the 50% and 95% HDRs of the forecast density. Note that the forecast density was bimodal for 1874 — hence the 50% HDR consisted of two sub-intervals for that year. The observed values are also given by the solid line.

Chapter 7

AN ALGORITHM FOR CONSTRUCTING HIGHEST DENSITY REGIONS

An algorithm is described which enables the rapid construction of highest density regions given a probability density function. The methods described are applicable to univariate data with a bounded, continuous density function containing any number of local modes. The density function may be known analytically or may be estimated from observations or simulations. A C version of the algorithm is also provided.

7.1 Description and Purpose

As defined in the previous chapter, a $100(1-\alpha)\%$ highest density region (HDR) is defined as:

1 the $100(1-\alpha)\%$ confidence region occupying the smallest possible volume in the sample space.

Equivalently, it can be defined as:

2 the $100(1-\alpha)\%$ confidence region in which every point has probability density at least as large as that of any point outside the region.

Such confidence regions are common in Bayesian analysis where they are applied using a posterior density (e.g. Berger (1985) or Box & Tiao (1973)). They are also variously known under other names such as "highest posterior density regions", "credible sets" and "Bayesian confidence intervals".

A HDR is of particular interest where the density contains several local modes or is difficult to handle analytically. In such cases, a HDR may appear unusual; for example, it could consist of several disjoint intervals. This often provides useful information which is 'masked' by other confidence regions.

It follows from definition **2** above, that a HDR is a set of the form $C(f^*) = \{x : f(x) \geq f^*\}$ where f^* is chosen so that the set has probability $1-\alpha$. In the univariate case, $C(f^*)$ can be written as the union of n disjoint sub-intervals,

$$C(f^*) = \bigcup_{j=1}^{n} (z_{2j-1}, z_{2j})$$

where $f(z_{2j-1}) = f(z_{2j}) = f^*, \quad j = 1, 2, \cdots, n$. Thus the problem is reduced to finding a suitable set of pairs $\{(z_{2j-1}, z_{2j}), j = 1, \cdots, n\}$. We will also be interested in calculating the probability of each sub-interval in the HDR. That is,

$$H_j = \int_{z_{2j-1}}^{z_{2j}} f(u)du \qquad j = 1, 2, \cdots, n. \tag{7.1.1}$$

Most work in the past on the calculation of HDRs has dealt with particular parametric forms of the density (see, for example, Lee, 1989). However, Wright

(1986) gives a procedure for numerical construction of HDRs where the density is univariate and unimodal, leading to a region consisting of a single interval. Berger (1985) gives the skeleton of a general algorithm for finding a HDR of a density $f(x)$ continuous in x.

The algorithm presented here is a development of Berger's algorithm. It enables rapid construction of a HDR for a given confidence level and a given univariate continuous probability density function, $f(x)$. The density may contain any number of local modes, and may be known analytically or may be estimated from observations or simulations. It is assumed that $f(x)$ is a bounded, continuous function of x and that the inverse of $f(x)$ is uniquely defined in the neighbourhood of each $f(z_j)$.

It is often very time consuming to calculate the density at a given point (e.g. with an empirical density). Hence, the algorithm requires the density to be calculated only on a grid of points $x_1 < x_2 < \cdots < x_N$. Elsewhere, interpolation is used to calculate the density. Both linear interpolation and cubic spline interpolation are implemented — the choice of method being specified at run time. Linear interpolation is most useful when $f(x)$ is not smooth. If cubic spline interpolation is used, then $f(x)$ must have continuous second order derivatives.

Densities with a finite number of discontinuities can be approximated by a continuous density with points on either side of each discontinuity, arbitrarily close together. Linear interpolation should be used in this case because the approximating density has discontinuous derivatives.

The end points, x_1 and x_N, should be chosen such that the required HDR has end points well within (x_1, x_N). In practice, I have found that $f(x_1)$ and $f(x_N)$ should be no more than about $0.9f^*$. This is because higher density regions are calculated in the algorithm while converging to the solution. Also, N should be large enough to provide sufficient information concerning the density. A value of $N = 50$ has been found satisfactory in practice.

7.2 Numerical Method

The algorithm may be outlined as follows:

Step 1 Choose some value f^*.

Step 2 Find all solutions to the equation $f(x) = f^*$ and construct the set $C(f^*) = \{x : f(x) \geq f^*\}$.

Step 3 Calculate $P(C(f^*)) = \int_{C(f^*)} f(u)du = \sum_{j=1}^{n} H_j$.

These steps should be repeated until a value of f^* is found such that $P(C(f^*)) = 1 - \alpha$ for some predetermined level α. Then the HDR is the set $C(f^*)$. Each step of the algorithm is now described in more detail.

Step 1

We wish to find f^* such that $g(f^*) = P(C(f^*)) - 1 + \alpha = 0$. The choice of f^* at each iteration is made using the Van Wijngaarden–Dekker–Brent method (also known as "Brent's algorithm") for finding roots which is described in Brent (1973) and Press et al. (1988). Given an interval in which a root is bracketed, this algorithm uses secant and false position techniques along with bisection

to enable superlinear convergence to the required root. In the accompanying code, $g(f^*)$ is calculated by the function `diffpcf()`.

Step 2

Step 2 can be processed rapidly by taking the ordered set of points $\{x_i\}$ at which the density is known, and finding pairs of points between which there exists a point with density f^*. That is, we wish to find pairs (x_{2i-1}, x_{2i}) such that $f(x_{2i-1}) < f^* < f(x_{2i})$ or $f(x_{2i}) < f^* < f(x_{2i-1})$. Then Brent's algorithm can be used again to find z_j such that $f(z_j) = f^*$ where $x_{2i-1} < z_j < x_{2i}$, using interpolation to calculate $f(x)$ where $x_{2i-1} < x < x_{2i}$. Provided the grid of points is sufficiently fine, all points at which the density equals f^* should be found.

This amounts to finding $f^{-1}(f^*)$. Therefore, it is assumed that $f^{-1}(\cdot)$ is uniquely defined in the neighbourhood of $f(z_j)$. If this is not true, (e.g. if $f(x)$ is constant in the neighbourhood of z_j) the algorithm may return a HDR with different confidence from that requested.

In the accompanying code, Step 2 is carried out by the function `invpdf()`.

Step 3

The integral in Step 3 is calculated using numerical integration. Again, to calculate the density at points other than those in the grid, interpolation is used.

Rather than recalculate the entire integral on each iteration, further improvements in speed can be achieved by simply calculating the *difference* between the integral $P(C(f^*))$ for the current iteration and the integral from the pre-

vious iteration. The new value of $P(C(f^*))$ is then obtained by adding this difference to the previous value of the integral. This optimisation for the integration is described in more detail in the following section.

Optimising the integration

Let t describe the iteration through steps 1–3, and

$$C_{t-1} = \bigcup_{k=1}^{n_{t-1}} (y_{2k-1}, y_{2k}), \qquad C_t = \bigcup_{j=1}^{n_t} (z_{2j-1}, z_{2j}),$$

$$H_{k,t-1} = \int_{y_{2k-1}}^{y_{2k}} f(u)du \qquad \text{and} \qquad H_{j,t} = \int_{z_{2j-1}}^{z_{2j}} f(u)du.$$

Here, C_{t-1} and C_t are successive approximations to the HDR. Assume $\{H_{k,t-1} : k = 1, \cdots, n_{t-1}\}$ are known (begin with $y_1 = y_2 =$ mode and $H_{1,0} = 0$). We want to find $\{H_{j,t} : j = 1, \cdots, n_t\}$ and

$$P(C_t) = \int_{C_t} f(u)du = \sum_{j=1}^{n} \int_{z_{2j-1}}^{z_{2j}} f(u)du = \sum_{j=1}^{n} H_{j,t}. \tag{7.2.1}$$

We wish to calculate $\{H_{j,t}\}$ from $\{H_{k,t-1}\}$ rather than calculate each $H_{j,t}$ directly from (7.1.1). To update the values of $\{H_{k,t-1}\}$, we integrate over the difference between $\{y_k\}$ and $\{z_j\}$.

Specifically, the sequence $\{H_{j,t}\}$ can be calculated as follows. For convenience, define $y_{2n_{t-1}+1}$ and z_{2n_t+1} to be $+\infty$. Consider each sub-interval (z_{2j-1}, z_{2j}) from C_t in turn, beginning with $j = 1$. At the same time, consider a sub-interval (y_{2k-1}, y_{2k}) from C_{t-1}, beginning with $k = 1$. Essentially, six situations can occur:

Figure 7.2.1: *Six possibilities when calculating successive approximations of an HDR*

1 $y_{2k-1} > z_{2j}$

This corresponds to Figure 7.2.1A. In this case,

$$H_{j,t} = \int_{z_{2j-1}}^{z_{2j}} f(u)du,$$

as there is no prior information in $\{H_{k,t-1}\}$ that is relevant to this sub-interval. Now move to the next sub-interval in C_t. That is, $j = j + 1$. But k remains unchanged.

2 $z_{2j-1} > y_{2k}$

This corresponds to Figure 7.2.1B. In this case, simply ignore $H_{k,t-1}$ and move to the next sub-interval in C_{t-1}. That is let $k = k + 1$; j remains unchanged.

3 $z_{2j} > y_{2(k+d)-1}, \quad d = 1, 2, \cdots.$

Figure 7.2.1C shows the case $d = 1$. $H_{j,t}$ can be calculated as follows.

$$H_{j,t} = \int_{z_{2j-1}}^{y_{2k-1}} f(u)du + \sum_{h=0}^{d-1} \left\{ H_{k+h,t-1} + \int_{y_{2(k+h)}}^{y_{2(k+h)+1}} f(u)du \right\} + H_{k+d,t-1} + \int_{y_{2(k+d)}}^{z_{2j}} f(u)du.$$

Also, $j = j + 1$ and $k = k + d + 1$.

4 $y_{2k} > z_{2(j+d)-1}, \quad d = 1, 2, \cdots.$

Figure 7.2.1D shows the case $d = 1$. In this case, it is easier to calculate each $H_{j,t}$ from (7.1.1). That is,

$$H_{j+h,t} = \int_{z_{2(j+h)-1}}^{z_{2(j+h)}} f(u)du.$$

where $h = 0, 1, \cdots, d$. Then $k = k + 1$ and $j = j + d + 1$.

5 $z_{2j-1} < y_{2k-1} < y_{2k} < z_{2j}$

This corresponds to Figure 7.2.1E. In this case,

$$H_{j,t} = H_{k,t-1} + \int_{z_{2j-1}}^{y_{2k-1}} f(u)du + \int_{y_{2k}}^{z_{2j}} f(u)du.$$

Also, $j = j + 1$ and $k = k + 1$.

6 $y_{2k-1} < z_{2j-1} < z_{2j} < y_{2k}$

This corresponds to Figure 7.2.1F. In this case,

$$H_{j,t} = H_{k,t-1} - \int_{y_{2k-1}}^{z_{2j-1}} f(u)du - \int_{z_{2j}}^{y_{2k}} f(u)du.$$

Also, $j = j + 1$ and $k = k + 1$.

Once $H_{j,t}$ has been calculated for $j = 1, \cdots, n$, $P(C_t)$ can be calculated from (7.2.1).

7.3 Structure

```
void hdr(float x[], float fx[], int capn, int interp,
          float *conf, float z[], int *n, float h[], int
*error);
```

Parameters:

`float x[]`	input:	contains the grid of values $x_1, \cdots, x_N$, at which the density has been calculated
`float fx[]`	input:	contains the density at each point in `x[]`
`int capn`	input:	N = the number of values in `x[]`

`int interp`	input:	indicates the method of interpolation. 1 = linear interpolation; 2 = cubic spline interpolation.
`float *conf`	input:	points to α = the required confidence for the HDR
	output:	points to the achieved probability of the HDR
`float z[]`	output:	contains the end points of the sub-intervals in the HDR, $z_1, \cdots, z_n$
`int *n`	output:	points to the number of sub-intervals in the HDR, n
`float h[]`	output:	contains the probability of each sub-interval in the HDR, $H_1, \cdots, H_n$
`int *error`	output:	points to an error indicator, equal to:

1: if capn < 2 (needed for interpolation)

2: if `conf` not between 0 and 1 or `interp` not equal to 1 or 2

3: if area under density less than `conf` or greater than 1

4: if can't find point at which density $= f^*$; longer tails are needed

5: if too many sub-intervals in HDR; increase value of MAXH.

0: otherwise.

It is possible that the achieved probability will be different from the requested probability. This will occur if one of the sub-intervals has an end point where the density is flat (so f^{-1} is not well defined). An example of this problem is given later.

Auxiliary routines

The following functions need to be provided:

1 `float brent(float (*func)(float), float a, float b,`
`float fa, float fb, float tol)`

This routine finds the root of a function `func` known to lie between `a` and `b`, using Brent's method. The argument `fa=func(a)` and `fb=func(b)`. The root, returned as `brent`, will be refined until its accuracy is `tol`.

The routine `zbrent` from Press et al. (1988, pp. 268–269), can be used with the following modifications:

- add the arguments `fa` and `fb`;
- delete the statement `fa=(*func)(a),fb=(*func)(b)`.

2 `void spline(float x[], float y[], int n, float yp1,`
`float ypn, float *y2)`

Given arrays `x[1..n]` and `y[1..n]` containing a tabulated function with `x[1]` $< \ldots <$ `x[n]` and given values `yp1` and `ypn` for the first derivative of the interpolating function at points 1 and n respectively, this routine returns an array `y2[1..n]` that contains the second derivatives of the interpolating function at the tabulated points `x`. If yp1 and/or ypn are equal to 1.0e30 or larger, the routine is signalled to set the corresponding boundary condition for a natural spline, with zero second derivative on that boundary.

The routine `spline` from Press et al. (1988, pp. 96–97) is suitable.

3 `void splint(float xa[], float ya[], float y2[], int n,`
`float x, float *y)`

Given the arrays `xa[1..n]` and `ya[1..n]`, which tabulate a function (with the `xa`'s in order) and given the array `y2[1..n]` which is the output from `spline()`, and given a value of `x`, this routine returns a cubic-spline interpolated value `y`.

The routine `splint` from Press et al. (1988, pp. 97) is suitable.

4 `float integral(float (*func)(float), float a, float b)`

Returns the integral of the function `func` from `a` to `b`. It should return zero if `a`=`b`. This is to cope with the cases in which a sub-interval has an end point at a near-singularity in the density.

Any good numerical integration procedure will suffice. For example, `qromb` from Press et al. (1988, p. 124), although this will need a simple modification so that zero is returned if `a`=`b`.

7.4 Testing and Timing

The time for a HDR to be constructed using the algorithm depends on the number of sub-intervals in the region and the slope of the density at the end points of those sub-intervals. Where the density at the end point of a sub-interval is nearly flat, the algorithm will take longer to construct the HDR because the value of H_j for that interval will be very sensitive to the position of the end points.

Two examples of densities were used to check the speed of the algorithm and

demonstrate the sources of variability in the speed. These were:

1 the standard normal p.d.f. calculated at evenly spaced points in $[-4, 4]$;

2 the empirical density of a data set consisting of 800 observations from a $N(10, 1)$ distribution and 200 observations from a $N(6, 1)$ distribution. The grid was calculated at evenly spaced points covering 1.5 times the range of the data.

Using an IBM compatible 386SX PC with a math co-processor, running at a landmark speed of 21 MHz, the following results were recorded. Times are CPU time for the routine `hdr()`. Spline interpolation was used.

$N = 50$	**Standard Normal P.D.F.**		**Mixture Distribution**	
Confidence	**Time**	**Interval**	**Time**	**Interval**
75%	0.22 secs	$[-1.150, 1, 150]$	0.82 secs	$[5.528, 5.722] \cup [8.039, 11.728]^*$
90%	0.20 secs	$[-1.645, 1.645]$	0.66 secs	$[5.074, 7.020] \cup [7.756, 11.923]$
95%	0.19 secs	$[-1.960, 1.960]$	0.41 secs	$[4.927, 12.035]$
99%	0.19 secs	$[-2.576, 2.576]$	0.39 secs	$[4.040, 12.670]$
$N = 100$				
75%	0.25 secs	$[-1.150, 1, 150]$	0.58 secs	$[5.551, 5.682] \cup [8.039, 11.726]$
90%	0.22 secs	$[-1.645, 1.645]$	0.60 secs	$[5.076, 7.023] \cup [7.758, 11.923]$
95%	0.22 secs	$[-1.960, 1.960]$	0.41 secs	$[4.927, 12.035]$
99%	0.22 secs	$[-2.576, 2.576]$	0.33 secs	$[4.039, 12.669]$

The interval marked * was returned with an achieved confidence of 75.48%. This was because both end points of the first sub-interval are near a local maximum in the density. The relatively long time taken to calculate the in-

Figure 7.4.1: *HDRs for the Mixture distribution based on a 100-point grid.*

terval reflects this. However, with the 100-point grid, this problem has been overcome.

Figure 7.4.1 shows the 75%, 90%, 95% and 99% HDRs for the mixture distribution based on a 100-point grid.

In general, the following heuristic rules (based on these tests and other experience) seem to apply:

- Where an end point of a sub-interval is near a point of inflection of the density, the algorithm may return a region of slightly different confidence from that requested. This is because the inverse of the density at that point is unstable. Increasing the number of points in the grid will often alleviate the problem. Alternatively, double precision rather than single

precision may be used.

- Densities with more than one local mode take longer than unimodal densities. In particular, when this leads to the HDR consisting of more than one interval, the time for calculation is increased.

- The number of points in the grid, N, has greatest effect with more complicated densities, affecting both the speed and accuracy of the results.

7.5 C code

```
/*****************************************************************************
 *****      ROUTINES FOR CALCULATING HIGHEST DENSITY REGIONS (HDRs)     *****
 *****                    (ANSI compatibility assumed)                  *****
 *****************************************************************************/

#include <stdlib.h>
#include <stdio.h>
#include <math.h>

#define TOLROOT       1.0e-5  /* Tolerance for root search               */
#define TOLAREA       1.0e-3  /* Tolerance for area under density        */
#define INFINITY      1.0e30  /* Value larger than any point in grid     */
#define MAXH              10  /* Maximum number of sub-intervals in HDR  */
#define TWOMAXH   (2 * MAXH)
#define TRUE               1
#define FALSE              0

/* ------------------------ Auxiliary Routines ----------------------------

    Four auxiliary routines need to be supplied by the user:
        brent(), spline(), splint() and integral().
    These routines may be taken from "Numerical Recipies in C" with little
    modification required.
*/

extern float integral(float (*)(float),float,float),
             brent(float (*)(float),float,float,float,float,float);
extern void  spline(float *, float *, int, float, float, float *),
             splint(float *, float *, float *, int, float, float *);


/* --------------------------- Sub-Routines ------------------------------ */

static float approxpdf(float),diffpcf(float),pdfminusf(float);
static int   invpdf(float *);
       void  lin_interp(float *, float *, int, float, float *);

/* -------------------------- Global Variables ----------------------------

    Most of these are made global to prevent additional arguments being
    passed to the auxiliary routines integral() and brent().
*/

float *gx,*gf,*gz,*gh,      /* Global pointers to xx[], fx[], z[] and h[]   */
      *gconf,               /* Global pointer to <conf>                     */
      fstar,                /* Value of f* for current approximation to HDR */
      *y2=NULL;             /* Second derivatives of density for use in
                               spline interpolation                         */
int   *gcapn,*gn,*gerror,   /* Global copies of <capn>, <n> and <error>     */
      ginterp,              /* Global copy of <interp>                      */
      imode,                /* Index to mode, i.e., xx[imode] is mode.
```

```
                                Needed in hdr() and diffpcf().  Made global to
                                prevent recalculation                          */
      newcalc;                /* Indicator for new calculation.  Used to check
                                if initialization necessary in diffpcf()     */


/* ---------------------------- Main routine ---------------------------- */

void hdr(float xx[], float fx[], int capn, int interp, float *conf,
            float z[], int *n, float h[], int *error)

/*
    Routine to calculate Highest Density Region for a given univariate
    density.  It is assumed that the density is a bounded continuous function
    and that the inverse of the density is well-defined near the end points of
    the sub-intervals of the HDR.

    Input required:
        It is assumed the density has been calculated at an ordered grid of
        points xx[0..capn-1], with the results stored in fx[0..capn-1].
        <interp> indicates the method of interpolation to use:
            interp = 1: linear interpolation
            interp = 2: cubic spline interpolation
        <conf> is the required confidence for the HDR.
    Results are returned in z[], <n>, h[], and <achieved>:
          z[0..2n-1]  contains the end points of the subintervals of the HDR;
          <n>         = the number of subintervals in the HDR;
          h[0..n-1]   contains the probability of each subinterval;
          <conf>      = the achieved probability of the HDR.
        It is assumed that z[] and h[] are sufficiently large to contain
        TWOMAXH and MAXH elements respectively.
    Error code is returned in <error>:
        0:  no error
        1:  capn < 2  (needed for interpolation)
        2:  <conf> not between 0 and 1 or interp not equal to 1 or 2.
        3:  area under specified density not between <conf> and 1.  Area is
            not required to equal one so long tails do not have to be given.
        4:  can't find point at which density = fstar; probably longer tails
            are needed
        5:  too many sub-intervals in HDR; increase value of MAXH.
    Restrictions:
        * The end points of the required HDR should be well within the range
          of xx[].  Provided this condition is met, the extreme tails of the
          distribution need not be specified.  If the condition is not met,
          error code 3 is returned.
        * Where the density is nearly constant at the end points of the
          sub-intervals in the HDR, the achieved confidence may differ
          slightly from the required confidence and it may take longer
          than otherwise to return a solution
    Method:
        Finds root of diffpcf() using brent().
*/
{
    int j;
    float area;
```

```
    /* Create global pointers to arguments for use in other functions */
    gx = xx; gf = fx; gz = z; gh = h;
    gcapn = &capn; gconf = conf; gn = n; gerror = error; ginterp = interp;

    /* Check input for errors 1, 2 and 3 */
    if(capn < 2)
        *error = 1;
    else if(*conf <= 0.0 || *conf >= 1.0 || interp < 1 || interp > 2)
        *error = 2;
    else
    {
        if(interp==2)
        {
            /* Create space for y2 and calculate second derivatives */
            if( (y2 = (float *)malloc((unsigned) capn*sizeof(float) )) == NULL )
            {
                printf("\n Insufficient memory for vector");
                exit(0);
            }
            spline(xx-1, fx-1, capn, 1.0e30, 1.0e30, y2-1);
        }
        area = integral(approxpdf,xx[0],xx[capn-1]);
        if(area > 1.0+TOLAREA || area < *conf)
            *error = 3;
        else
            *error = 0;
    }

    if(*error==0)
    {
        newcalc = TRUE;
        /* Find index of mode */
        imode = capn/2;
        for(j=0;j<capn;j++)
        {
            if(fx[j]>fx[imode])
                imode = j;
        }

        /*** Main statement: search for root ***/
        brent(diffpcf,0.0,fx[imode],1.0-(*conf),-(*conf),fx[imode]*TOLROOT);
    }

    if(interp==2 && y2)
        free(y2);

    if(*error==0)
    {
        /* Calculate achieved confidence */
        *conf = 0.0;
        for(j=0;j<*n/2;j++)
            *conf += h[j];
    }
}
```

```
/* -------------------------- Sub-routines ------------------------------ */

static float diffpcf(float newfstar)

/*  Sub-routine returns the difference between the required confidence and
    the confidence of the current estimate of the HDR.  When this routine
    returns a value sufficiently close to zero, we have found the required
    HDR.  Most of the routine consists of code implementing the optimisation
    of the integration.

    Input: <newfstar> = new estimate of f*.

    Called by: brent() which in turn is called by hdr().

    Uses: invpdf() to find inverse of p.d.f. at <newfstar>
          integral() to do numerical integration.

    If an error occurs, it is recorded in <gerror> and the routine returns
    zero to force brent() to return to hdr()
*/

{
    int currhump,            /* Current subscript of gh = (j+1)/2         */
        lasthump,            /* Subscript of last value in gh             */
        j,k,l;
    float pcf;               /* Probability of current approx. to HDR     */
    static float y[TWOMAXH]; /* Copy of gz from previous iteration        */
    static int m;            /* Number of points in y[]                   */

    /* Initialize y[] */
    if(newcalc)
    {
        /* First approximation to HDR is zero interval with end points equal
            to the mode */
        y[0] = y[1] = gx[imode];
        y[2] = INFINITY;
        m=2;
        gh[0] = 0.0;
        newcalc = FALSE;
    }

    /* Calculate intervals in new HDR estimate */
    fstar = newfstar;
    *gn = invpdf(gz);

    /* Check that all intercepts found */
    if(*gerror == 5)
        return(0.0);
    else if(*gn==0 || *gn%2==1)
    {
        *gerror = 4;
        return(0.0);
    }

    /* Calculate integral of density over current HDR estimate.
```

```
   Implements optimisation techniques by integrating only over the
   difference between the previous and the current HDR estimate */

lasthump = max(*gn,m)/2;
k = currhump = 0;
for(j=0;j<*gn;)
{
    if(gz[j+1] < y[k])
    {
        /* Case 1: add new hump */
        for(l=lasthump;l>currhump;l--)
            gh[l] = gh[l-1];   /* make room in gh */
        gh[currhump] = integral(approxpdf,gz[j],gz[j+1]);
        j += 2;
        currhump++;
    }
    else if(y[k+1] < gz[j])
    {
        /* Case 2: remove old hump */
        for(l=currhump;l<lasthump;l++)
            gh[l] = gh[l+1];  /* remove room in gh */
        k += 2;
    }
    else if(gz[j+1] > y[k+2])
    {
        /* Case 3: several humps become 1 - include valleys */
        gh[currhump] += integral(approxpdf,gz[j],y[k]);
        k += 2;
        do{
            gh[currhump] += integral(approxpdf,y[k-1],y[k]);
            k += 2;
            gh[currhump] += gh[currhump+1];
            for(l=currhump+1;l<lasthump;l++)
                gh[l] = gh[l+1]; /* remove hump */
        }while(gz[j+1] > y[k] && k<=m);
        gh[currhump] += integral(approxpdf,y[k-1],gz[j+1]);
        j += 2;
        currhump++;
    }
    else if(y[k+1] > gz[j+2])
    {
        /* Case 4: 1 hump becomes several - recalculate area under
                                                          each hump */
        do{
            gh[currhump] = integral(approxpdf,gz[j],gz[j+1]);
            j += 2;
            currhump++;
            for(l=lasthump;l>currhump;l--)
                gh[l] = gh[l-1];   /* make room in gh */
        }while(y[k+1] > gz[j] && j<=*gn);
        k += 2;
    }
    else
    {
        /* Case 5 & 6: same hump */
```

```
            gh[currhump] += integral(approxpdf,gz[j],y[k])
                                    + integral(approxpdf,y[k+1],gz[j+1]);
            k += 2;
            j += 2;
            currhump++;
        }
    }

    /* Add up area under humps */
    pcf = gh[0];
    for(j=1;j<*gn/2;j++)
        pcf += gh[j];

    /* Save new estimate of HDR for next iteration */
    for(j=0;j<=*gn;j++)
        y[j] = gz[j];
    m = *gn;

    return(pcf - *gconf);
}

/* ------------------------------------------------------------------------ */

static float approxpdf(float u)

/*  Sub-routine returns the value of the density at <u>.  It uses cubic
    spline interpolation between values in the grid supplied.

    Called by: integral() and pdfminusf().

    Uses: Either splint() or lin_interp() to do interpolation.

    N.B. array arguments to splint() are decremented as in
        "Numerical Recipies in C"
*/
{
    float midx;

    if(ginterp == 1)
        lin_interp(gx, gf, *gcapn, u, &midx);
    else
        splint(gx-1, gf-1, y2-1, *gcapn, u, &midx);

    return(midx);
}

/* ------------------------------------------------------------------------ */

static float pdfminusf(float u)

/*  Sub-routine returns the value of the density at <u> minus the current
    value of f*.  This is used by invpdf() to find the inverse p.d.f.
    at f*.

    Called by: brent() which in turn is called by invpdf().
```

```
*/
{
    return(approxpdf(u)-fstar);
}

/* ------------------------------------------------------------------------ */

static int invpdf(float *gz)

/*  Sub-routine to calculate the inverse p.d.f. at current value of f*.
    Finds all points at which f(x)=fstar and saves them in gz.  Function
    returns the number of points in gz.

    Called by: diffpcf()

    Uses brent() to find roots of pdfminusf().
*/
{
    int i,j;
    float dif1,dif2;

    /* Loop over all points in grid looking for points that bracket an
       intercept */
    j=0;
    for(i=0;i<*gcapn-1 && j<TWOMAXH;i++)
    {
        dif1=gf[i]-fstar;
        if(dif1==0.0)
            gz[j++] = gx[i];
        else
        {
            dif2=gf[i+1]-fstar;
            if(dif1*dif2<0.0)
                /* Find root of pdfminusf that lies between gx[i] and gx[i+1] */
                gz[j++] = brent(pdfminusf, gx[i], gx[i+1], dif1, dif2,
                                fabs(gx[i]*TOLROOT));
        }
    }
    if(j>=TWOMAXH)
        *gerror = 5;
    else
        gz[j]=INFINITY;
    return(j);
}

/* ------------------------------------------------------------------------ */

void lin_interp(float *x, float *y, int n, float xx, float *yy)

/* Given arrays x[0..n-1] and y[0..n-1], and given value xx, this routine
    returns a value yy = P(xx) where P(xx) is a linear interpolating
    function.
*/

{
```

```
    int klo, khi, k;
    float m, h;

    /* find place in table using bisection */
    klo = 0;
    khi = n-1;
    while(khi-klo > 1)
    {
        k = (khi+klo) >> 1;
        if(x[k] > xx)
            khi = k;
        else
            klo = k;
    }

    h = x[khi] - x[klo];
    if(h==0.0F)
    {
        printf("\n Bad XA input to routine lin_interp()");
        exit(0);
    }
    m = (y[khi] - y[klo]) / h;
    *yy = m*(xx-x[klo]) + y[klo];
}
```

APPENDIX: STEEL FURNACE TEMPERATURE DATA

Data are temperatures (in degrees Celsius) from the top of a steel blast furnace. The first measurement was taken at 12.06 am, 7 August 1990, and subsequent measurements were made every 6 minutes. The last measurement was taken at 11.54 pm, 16 August 1990 making 2399 observations. Ten observations are missing and marked with an asterisk (*). Data are to be read across the rows from left to right.

7 August 1990										
	0 min	6 min	12 min	18 min	24 min	30 min	36 min	42 min	48 min	54 min
0 hr		141.43	144.18	156.89	154.42	164.33	157.90	151.10	158.57	141.56
1 hr	155.06	170.67	169.31	160.16	166.99	173.66	156.21	164.13	159.42	169.51
2 hr	143.37	161.84	148.13	165.68	174.00	160.34	173.25	161.81	146.84	160.66
3 hr	151.06	142.37	155.26	144.27	155.39	150.25	127.47	153.23	153.61	138.21
4 hr	141.57	159.39	147.40	139.50	148.45	146.86	156.86	140.99	127.19	124.42
5 hr	128.84	141.69	128.35	126.41	132.69	130.67	121.21	129.89	115.23	119.25
6 hr	138.35	128.31	132.73	134.49	145.25	152.62	140.06	139.89	152.75	175.77
7 hr	152.71	147.48	150.40	134.04	122.09	126.96	116.22	120.62	130.60	123.48
8 hr	131.71	122.16	118.26	128.92	125.32	128.10	169.53	140.21	148.48	133.20
9 hr	160.21	145.84	157.27	158.21	182.03	153.13	141.13	143.86	135.07	127.98
10 hr	124.78	*	116.87	*	125.36	136.85	151.72	140.68	143.67	142.02
11 hr	135.14	161.45	145.99	140.37	140.67	141.14	142.22	130.22	154.79	132.77
12 hr	147.49	137.38	122.56	117.99	117.77	131.69	127.41	144.97	133.93	143.86
13 hr	126.45	127.30	146.79	131.08	125.16	134.97	130.37	137.34	134.98	138.61
14 hr	132.95	143.67	150.12	152.67	132.73	132.32	145.63	134.50	129.64	138.99
15 hr	129.63	128.30	148.97	147.80	141.75	133.60	152.20	141.47	151.84	154.85
16 hr	151.08	143.40	145.17	159.72	143.23	143.21	156.63	148.50	169.48	155.32
17 hr	161.88	135.80	134.26	142.71	141.26	153.75	154.86	148.31	147.21	143.10
18 hr	139.83	147.53	150.80	159.21	148.48	147.44	140.52	159.10	143.78	134.82
19 hr	135.28	131.42	136.75	140.50	130.53	147.50	139.77	166.60	152.06	153.29
20 hr	140.67	129.26	128.09	133.69	159.09	136.90	126.92	132.08	135.99	134.76
21 hr	150.97	172.79	151.39	146.38	166.95	145.79	131.32	146.36	177.96	151.02
22 hr	148.65	165.56	178.67	169.91	150.04	144.95	158.50	152.58	164.77	156.17
23 hr	158.08	144.04	164.86	159.21	160.10	155.02	147.77	164.17	146.63	164.48

8 August 1990										
	0 min	6 min	12 min	18 min	24 min	30 min	36 min	42 min	48 min	54 min
0 hr	157.31	139.79	154.74	159.86	149.92	146.59	144.37	154.85	149.74	165.32
1 hr	148.25	171.47	171.18	181.76	162.41	161.20	160.10	178.30	160.61	159.02
2 hr	171.63	159.54	158.54	158.19	142.68	147.73	145.93	160.06	146.64	150.35
3 hr	152.12	147.09	135.94	139.99	130.73	135.35	135.94	155.28	140.52	152.04
4 hr	144.49	153.30	154.92	142.39	163.43	179.65	164.75	172.96	166.26	160.04
5 hr	156.63	143.74	156.98	154.68	150.28	143.92	136.71	142.87	147.02	146.90
6 hr	148.83	148.22	138.35	149.46	171.57	159.31	144.58	166.52	165.62	152.58
7 hr	143.89	162.57	149.20	153.73	161.03	142.82	133.70	141.68	158.68	140.68
8 hr	137.62	155.40	148.30	133.43	139.71	139.05	130.07	125.72	125.82	118.37
9 hr	119.18	146.38	133.56	131.78	130.06	131.53	128.76	125.32	152.05	140.93
10 hr	113.91	119.67	126.38	136.58	138.00	152.65	165.59	168.61	168.09	174.06
11 hr	167.18	161.23	172.98	155.01	164.87	177.88	173.70	170.10	164.18	181.61
12 hr	167.80	170.90	165.30	192.31	173.60	167.85	176.59	185.32	173.35	195.18
13 hr	182.06	177.35	160.97	143.64	153.74	144.29	137.61	143.13	139.00	148.66
14 hr	137.78	127.84	132.18	129.41	150.86	141.45	135.34	141.26	162.07	143.24
15 hr	152.67	147.84	150.01	135.74	125.19	*	119.67	114.79	130.91	118.69
16 hr	118.51	*	124.64	142.93	148.41	143.60	134.34	145.20	139.52	137.36
17 hr	152.33	160.77	159.39	154.91	154.61	148.26	160.72	150.78	164.55	153.34
18 hr	145.89	154.15	172.84	164.88	164.90	165.61	156.80	148.17	143.63	154.47
19 hr	149.07	146.11	162.54	177.60	172.33	157.43	172.28	168.34	190.06	154.45
20 hr	147.89	146.62	157.90	145.08	143.80	163.81	152.02	144.24	154.15	157.59
21 hr	144.06	138.68	141.59	139.27	141.36	152.59	138.46	134.35	146.57	143.07
22 hr	144.93	139.67	155.84	144.61	138.76	148.99	151.59	139.65	141.64	146.32
23 hr	139.49	129.61	139.81	150.14	128.97	121.25	133.83	131.96	123.14	128.17

9 August 1990										
	0 min	6 min	12 min	18 min	24 min	30 min	36 min	42 min	48 min	54 min
0 hr	124.04	110.72	125.92	119.41	114.80	127.60	*	124.56	*	116.50
1 hr	107.63	129.65	124.14	121.70	116.80	120.14	131.52	129.87	122.02	139.26
2 hr	147.15	135.58	134.34	130.60	122.40	149.10	134.37	149.58	137.76	125.70
3 hr	140.43	150.42	131.14	131.46	134.28	132.56	131.13	134.70	131.96	122.23
4 hr	123.43	158.44	145.03	139.89	138.78	145.79	133.94	117.14	131.46	133.11
5 hr	128.40	127.88	130.91	136.62	116.50	112.49	113.04	121.21	117.25	111.39
6 hr	128.16	108.62	107.53	132.41	125.63	139.62	136.44	163.56	170.92	165.76
7 hr	155.61	153.34	186.80	173.83	151.99	154.72	180.97	163.75	145.80	135.58
8 hr	142.67	126.95	119.44	133.37	128.39	126.38	136.22	130.31	136.76	134.15
9 hr	135.51	134.81	131.97	154.44	144.23	133.14	138.06	141.97	134.87	107.08
10 hr	103.25	101.69	106.36	116.33	120.97	110.43	122.81	124.85	115.81	127.22
11 hr	117.88	133.20	121.59	112.64	116.84	116.69	135.04	154.14	134.82	138.26
12 hr	143.15	150.13	129.65	123.53	147.47	148.20	139.21	170.02	163.50	189.20
13 hr	158.62	163.03	162.88	160.25	157.96	148.95	156.36	129.98	120.39	135.92
14 hr	136.46	110.14	112.70	146.54	144.14	135.14	143.32	154.90	142.88	128.95
15 hr	153.55	151.68	136.50	140.98	144.37	141.64	132.39	147.95	138.37	137.16
16 hr	115.84	136.73	128.30	134.18	111.88	105.83	121.19	120.72	122.72	128.66
17 hr	124.62	151.07	164.83	147.59	197.58	171.68	160.85	157.13	161.69	181.30
18 hr	192.79	162.70	148.62	171.99	153.24	179.05	164.51	201.93	190.68	186.76
19 hr	170.85	156.03	150.52	151.02	139.04	128.69	137.58	151.72	133.82	125.70
20 hr	143.76	145.73	134.66	140.95	168.40	151.90	144.77	152.51	168.05	170.31
21 hr	172.45	167.19	148.48	167.12	147.00	159.18	146.60	143.00	130.34	130.43
22 hr	141.89	120.89	121.20	153.62	149.26	148.42	151.96	173.82	155.83	166.66
23 hr	167.58	166.63	182.58	172.09	175.62	170.19	193.29	170.16	162.59	165.21

10 August 1990										
	0 min	6 min	12 min	18 min	24 min	30 min	36 min	42 min	48 min	54 min
0 hr	174.88	159.47	152.15	184.64	175.67	158.60	164.05	155.33	149.20	148.64
1 hr	174.23	161.69	147.30	156.18	146.95	143.75	144.10	148.20	155.22	152.29
2 hr	159.75	155.31	160.70	160.19	181.20	175.35	156.32	169.27	186.98	178.41
3 hr	188.19	183.50	173.15	180.59	161.57	176.00	170.72	177.96	165.56	184.19
4 hr	170.88	165.61	161.75	166.30	168.52	153.65	180.18	173.36	169.83	161.33
5 hr	171.29	154.80	139.12	133.61	139.20	126.55	138.68	125.79	115.10	130.26
6 hr	131.08	128.90	132.15	134.72	140.87	141.68	124.99	129.36	149.90	132.02
7 hr	120.69	140.43	138.22	124.33	130.33	132.73	123.16	116.56	111.40	117.33
8 hr	109.45	118.00	116.18	114.41	114.36	113.16	107.01	105.74	102.12	102.90
9 hr	106.12	125.38	102.31	93.72	106.32	135.89	111.37	114.69	123.35	106.82
10 hr	108.95	110.52	106.11	107.08	*	118.55	111.58	109.68	126.65	117.77
11 hr	121.97	119.81	116.53	115.30	107.82	112.21	113.14	122.20	105.25	114.05
12 hr	118.02	115.27	119.97	111.16	109.95	118.91	125.66	131.61	135.40	144.25
13 hr	137.02	139.92	128.87	152.83	146.71	159.66	157.11	169.22	161.91	168.25
14 hr	157.03	168.81	165.49	167.38	169.07	160.97	159.72	144.73	136.92	131.48
15 hr	127.28	137.74	129.43	112.25	126.41	116.04	117.84	110.50	141.56	123.85
16 hr	115.30	106.04	144.74	125.79	105.26	107.95	142.03	117.05	128.29	109.38
17 hr	126.40	109.32	100.21	112.55	114.29	124.51	111.80	108.40	138.38	120.68
18 hr	121.04	139.18	170.98	165.10	181.29	159.61	202.43	239.44	193.11	159.59
19 hr	164.38	155.19	153.16	153.50	146.36	137.45	127.43	120.56	128.24	128.86
20 hr	161.81	145.07	175.54	145.75	145.35	169.45	137.15	117.22	119.14	107.78
21 hr	104.76	135.72	128.83	120.12	130.72	126.45	136.85	150.50	193.23	175.96
22 hr	161.12	150.41	134.96	134.91	130.75	134.10	157.02	134.56	125.19	117.55
23 hr	122.20	111.19	107.12	118.20	108.74	105.41	115.24	113.77	138.92	117.88

11 August 1990										
	0 min	6 min	12 min	18 min	24 min	30 min	36 min	42 min	48 min	54 min
0 hr	108.90	118.72	104.13	118.57	119.36	148.37	138.03	128.93	128.41	166.99
1 hr	142.19	132.28	145.90	185.85	169.33	140.86	160.35	164.55	137.44	143.10
2 hr	165.32	150.37	145.25	159.22	156.30	164.03	159.09	144.29	128.53	149.30
3 hr	139.83	159.49	150.74	185.33	161.96	166.08	139.39	140.03	153.38	189.84
4 hr	148.84	136.01	159.08	134.29	124.90	135.67	144.23	147.27	141.99	154.51
5 hr	148.40	147.66	163.54	164.65	157.67	152.40	175.18	158.60	154.82	139.83
6 hr	182.31	173.97	170.82	154.55	166.07	159.08	152.48	158.46	160.89	173.40
7 hr	169.93	147.98	166.43	152.83	138.30	139.33	162.88	143.99	123.52	137.33
8 hr	144.23	163.65	170.06	141.89	130.41	143.30	120.48	162.33	137.94	122.01
9 hr	121.58	120.16	130.00	161.92	132.13	135.01	123.55	122.15	143.01	125.52
10 hr	110.34	112.73	117.25	145.40	132.95	126.29	131.62	173.33	144.50	125.13
11 hr	154.52	138.81	119.15	140.57	156.15	139.93	124.38	146.83	156.54	139.23
12 hr	131.07	165.48	147.12	138.97	154.12	189.24	162.89	210.83	192.26	197.03
13 hr	176.30	192.70	170.39	157.79	147.92	146.90	152.83	131.82	128.77	146.99
14 hr	130.67	137.66	155.15	126.47	125.65	143.25	126.70	124.29	144.57	134.93
15 hr	145.53	143.78	165.91	153.71	155.36	162.01	184.29	159.78	164.22	163.33
16 hr	188.72	167.60	159.58	184.25	175.00	166.70	156.76	193.94	174.96	175.42
17 hr	160.52	176.83	153.64	148.37	166.88	140.04	133.81	150.04	155.35	144.91
18 hr	140.69	162.45	135.21	135.27	159.25	147.93	140.64	141.12	148.16	138.41
19 hr	139.55	162.92	150.95	158.57	140.12	170.09	151.42	184.24	164.09	168.49
20 hr	160.97	172.20	156.33	161.30	147.49	135.43	166.27	173.94	171.51	154.07
21 hr	132.63	127.44	125.21	114.91	125.80	119.55	108.16	124.81	118.65	123.26
22 hr	114.76	121.68	113.06	112.21	142.19	135.28	141.37	149.38	188.51	166.99
23 hr	165.14	147.31	193.83	166.79	149.92	158.73	162.61	178.04	179.07	164.42

12 August 1990										
	0 min	6 min	12 min	18 min	24 min	30 min	36 min	42 min	48 min	54 min
0 hr	176.10	161.89	170.50	164.94	199.18	179.58	174.10	161.72	170.61	155.98
1 hr	134.80	127.77	128.53	116.22	120.84	120.48	114.82	159.22	150.23	146.82
2 hr	141.63	153.50	153.33	157.21	143.36	155.94	163.71	155.37	137.88	157.82
3 hr	144.91	133.34	145.13	130.33	117.50	128.06	135.58	143.11	128.80	148.52
4 hr	160.24	144.51	140.92	150.03	136.81	129.76	152.53	144.94	143.76	137.91
5 hr	196.23	167.75	152.04	150.25	146.85	153.77	141.55	158.51	151.31	160.58
6 hr	153.49	174.09	169.00	153.37	157.41	153.72	168.32	157.04	184.82	172.50
7 hr	170.83	163.97	159.04	157.66	137.78	145.32	140.50	129.06	147.53	125.75
8 hr	138.98	155.79	150.12	136.85	128.82	172.89	149.72	126.85	126.90	118.88
9 hr	125.17	138.53	141.59	147.05	145.58	152.81	161.59	143.59	136.92	155.44
10 hr	161.38	165.05	177.16	169.33	156.60	142.96	133.87	162.05	160.55	150.43
11 hr	154.63	169.42	155.57	147.90	169.22	157.39	151.71	152.12	188.55	179.47
12 hr	168.92	165.99	176.54	159.51	144.53	169.10	206.08	195.53	283.00	372.75
13 hr	433.59	403.91	299.06	267.73	272.57	353.79	299.33	253.79	219.00	183.89
14 hr	151.87	132.59	123.03	114.42	108.48	102.31	100.57	101.85	95.77	97.52
15 hr	100.04	93.71	93.36	93.94	91.24	89.85	97.64	100.39	100.90	95.17
16 hr	103.34	104.73	114.96	108.46	121.42	118.80	123.21	*	131.61	117.12
17 hr	126.31	121.55	115.65	118.40	114.35	134.97	125.25	133.15	124.77	133.63
18 hr	126.51	117.05	137.86	136.20	152.33	136.98	138.77	132.16	128.55	140.16
19 hr	129.32	122.48	136.01	163.81	147.85	143.55	156.12	162.09	154.42	154.91
20 hr	167.03	166.25	147.92	140.59	169.05	148.37	154.23	158.65	175.43	149.59
21 hr	140.09	131.39	122.93	120.35	119.43	123.01	126.55	132.41	128.51	123.72
22 hr	134.10	130.78	143.21	141.47	150.22	143.63	128.52	149.14	171.30	152.67
23 hr	144.18	155.62	151.59	142.40	146.38	183.31	173.20	160.61	169.69	227.55

13 August 1990										
	0 min	6 min	12 min	18 min	24 min	30 min	36 min	42 min	48 min	54 min
0 hr	227.96	179.21	154.01	137.23	131.46	127.16	130.15	124.30	143.59	133.44
1 hr	142.63	210.51	281.55	328.68	265.41	202.59	160.48	138.35	127.78	119.50
2 hr	111.51	103.96	100.49	108.57	111.88	114.75	120.29	130.02	126.11	119.52
3 hr	124.99	126.22	139.52	128.15	116.32	132.40	128.03	128.36	125.53	150.61
4 hr	134.68	131.92	150.95	138.13	158.47	213.27	257.23	205.36	166.71	145.38
5 hr	135.12	129.04	120.62	121.73	116.29	110.69	111.14	108.43	102.84	98.80
6 hr	97.38	99.60	100.12	96.03	96.40	99.66	89.60	87.69	90.48	88.35
7 hr	86.13	93.22	91.54	92.61	100.92	98.81	95.09	102.27	120.08	111.85
8 hr	111.61	110.32	107.20	105.61	99.27	100.09	110.75	104.23	115.32	109.51
9 hr	112.16	111.88	114.02	121.79	124.39	123.44	122.72	134.74	125.38	135.85
10 hr	132.77	141.26	122.85	134.06	126.70	119.72	141.71	137.63	146.09	190.05
11 hr	211.78	169.91	146.56	137.92	147.29	143.61	147.17	147.86	166.65	153.24
12 hr	147.68	142.75	165.44	155.54	157.18	152.72	170.04	151.75	145.07	160.61
13 hr	145.35	150.49	149.22	151.47	152.21	153.90	160.31	138.43	132.32	157.99
14 hr	133.00	126.58	142.79	128.99	121.72	136.74	137.95	131.55	128.02	152.71
15 hr	142.62	135.27	140.13	129.40	129.01	135.71	131.35	126.20	131.04	153.62
16 hr	147.05	145.95	143.89	157.98	144.93	157.58	146.42	136.88	145.49	137.94
17 hr	127.20	145.67	139.35	148.94	144.45	134.88	142.87	130.90	139.13	133.04
18 hr	147.01	138.23	138.47	139.07	146.82	142.39	134.54	153.97	144.47	161.58
19 hr	150.08	147.24	144.99	197.42	182.22	189.92	169.12	151.93	149.93	148.59
20 hr	143.98	139.76	132.30	128.92	131.85	136.89	131.88	136.15	132.31	139.10
21 hr	141.20	148.82	138.91	130.40	143.68	145.24	156.57	143.63	162.54	154.21
22 hr	153.76	161.04	162.86	159.01	152.04	163.18	154.49	169.30	148.02	165.65
23 hr	158.26	150.10	158.87	170.37	160.31	158.06	161.22	155.39	149.68	152.17

14 August 1990										
	0 min	6 min	12 min	18 min	24 min	30 min	36 min	42 min	48 min	54 min
0 hr	161.14	148.04	139.63	160.68	145.16	145.07	137.69	130.38	149.19	144.75
1 hr	151.20	136.83	123.51	151.44	175.99	159.28	150.90	141.62	134.94	177.30
2 hr	149.01	154.77	154.22	131.08	166.02	174.99	168.43	146.53	177.89	199.35
3 hr	173.25	215.91	168.04	161.47	164.07	159.02	189.80	169.39	149.82	150.17
4 hr	142.18	136.77	140.84	132.10	130.62	141.22	133.98	130.39	142.43	127.43
5 hr	135.35	148.92	132.98	127.25	135.80	135.26	124.29	120.96	130.13	123.99
6 hr	122.55	127.08	121.28	112.87	132.68	133.96	117.98	112.21	103.06	104.52
7 hr	107.06	127.71	109.76	110.66	104.02	103.52	100.21	89.97	109.82	112.75
8 hr	117.34	117.60	176.88	167.53	150.72	131.52	121.29	121.78	114.53	110.46
9 hr	120.74	111.58	133.51	119.09	117.97	123.26	140.76	135.54	122.72	127.67
10 hr	124.71	161.68	143.41	145.80	149.51	157.89	150.11	131.45	132.19	142.30
11 hr	144.05	148.26	191.60	162.70	216.53	300.97	367.99	391.78	409.91	420.62
12 hr	415.96	316.11	236.84	181.65	156.71	139.16	127.24	121.20	116.93	109.19
13 hr	109.96	116.03	116.06	113.97	119.60	110.13	109.17	109.61	106.67	103.02
14 hr	101.57	99.58	102.34	109.70	105.58	116.22	111.61	111.04	110.27	107.42
15 hr	106.70	116.45	135.83	121.31	119.35	123.21	167.98	228.57	284.41	324.07
16 hr	236.32	173.21	144.83	124.55	111.71	106.65	112.90	121.69	131.68	128.89
17 hr	123.94	132.61	118.13	106.86	120.91	118.55	128.83	112.67	113.65	113.50
18 hr	135.62	121.74	119.38	124.41	122.42	116.01	123.52	136.22	132.30	120.31
19 hr	137.99	136.08	130.56	129.76	119.65	127.07	125.78	124.91	118.77	133.46
20 hr	140.17	124.30	112.06	127.73	148.66	123.32	132.27	149.26	138.09	139.18
21 hr	146.50	134.99	126.22	134.20	150.83	144.24	156.22	155.65	147.41	154.17
22 hr	143.02	153.86	145.83	157.67	138.81	134.27	138.13	135.50	131.59	151.00
23 hr	178.13	159.21	163.46	168.41	189.64	162.53	152.14	177.47	156.19	155.84

15 August 1990										
	0 min	6 min	12 min	18 min	24 min	30 min	36 min	42 min	48 min	54 min
0 hr	172.40	184.65	160.12	151.48	181.23	157.21	144.05	154.35	145.18	155.87
1 hr	153.06	154.22	145.44	156.15	156.34	141.95	148.07	138.38	149.20	143.70
2 hr	158.35	157.87	159.26	155.37	145.52	166.75	168.94	158.83	164.13	156.69
3 hr	152.19	151.80	178.72	159.26	161.68	159.61	211.46	287.30	245.47	205.00
4 hr	175.29	160.23	155.79	149.76	143.54	152.25	149.18	137.46	126.47	141.35
5 hr	148.23	155.90	171.21	162.42	172.70	171.30	152.19	155.97	160.14	199.13
6 hr	175.34	152.42	150.38	145.60	164.30	154.79	157.94	147.73	163.45	155.47
7 hr	150.23	146.34	142.09	144.52	138.45	141.09	157.12	163.64	147.37	149.71
8 hr	152.59	160.53	136.29	133.16	150.03	164.98	138.52	139.05	145.06	134.27
9 hr	124.46	144.89	150.05	133.92	131.03	161.86	142.57	138.00	147.09	163.84
10 hr	148.24	132.86	163.83	151.46	149.62	147.67	144.88	146.50	138.67	154.63
11 hr	143.38	161.94	150.74	148.52	*	144.49	169.95	151.43	140.86	151.51
12 hr	148.57	153.47	158.53	165.38	169.65	157.87	158.28	147.49	159.74	154.42
13 hr	143.25	160.20	150.90	144.78	146.60	161.32	153.09	143.74	153.23	150.86
14 hr	154.96	140.06	143.63	142.09	144.63	164.96	236.62	312.35	338.28	241.63
15 hr	264.61	248.53	218.20	175.87	164.02	151.78	134.48	129.87	132.03	128.29
16 hr	123.51	123.10	125.95	*	117.28	111.84	105.60	105.90	98.88	98.76
17 hr	98.68	91.64	89.30	88.84	86.32	85.07	87.77	96.55	91.79	96.91
18 hr	97.38	100.19	97.72	92.22	93.11	86.13	84.53	83.36	81.14	84.36
19 hr	81.45	75.86	74.23	71.40	71.15	71.38	70.13	69.76	69.91	70.85
20 hr	71.38	74.03	75.88	75.13	73.91	74.05	73.42	74.52	75.13	77.86
21 hr	77.07	77.47	78.88	79.93	81.74	80.24	81.25	79.97	77.65	80.24
22 hr	79.49	80.13	82.97	83.79	82.30	81.56	83.41	86.60	85.21	88.12
23 hr	90.46	92.60	93.99	93.97	94.36	93.35	99.75	97.08	97.26	107.94

16 August 1990										
	0 min	6 min	12 min	18 min	24 min	30 min	36 min	42 min	48 min	54 min
0 hr	96.33	110.81	104.72	101.38	105.00	94.89	104.80	101.87	120.08	115.71
1 hr	110.16	109.82	108.97	99.99	101.10	101.47	94.61	103.09	99.16	96.06
2 hr	105.46	108.68	108.86	101.53	114.08	104.14	95.97	104.19	98.25	94.15
3 hr	98.48	112.44	91.06	91.21	95.44	83.67	89.21	89.44	85.21	84.25
4 hr	94.50	90.99	91.14	89.97	101.90	91.60	91.34	95.47	93.11	104.32
5 hr	107.65	105.85	107.15	104.68	106.37	111.03	107.92	113.62	124.85	128.61
6 hr	113.06	112.15	128.32	148.56	137.60	131.14	139.99	153.50	155.93	149.47
7 hr	144.33	133.17	157.10	147.06	132.83	134.08	134.39	126.49	127.87	137.38
8 hr	115.04	117.54	118.84	148.08	119.48	104.12	114.95	110.35	130.26	120.57
9 hr	107.95	116.21	110.05	135.97	126.86	117.03	125.49	116.50	115.77	109.27
10 hr	101.43	108.14	102.84	95.82	103.68	98.11	97.30	102.41	102.67	110.90
11 hr	105.83	110.07	105.31	117.86	112.56	135.39	140.83	125.08	117.42	112.07
12 hr	118.72	117.45	130.51	123.30	132.27	122.63	136.62	131.56	124.55	126.89
13 hr	150.53	136.58	158.80	144.91	156.67	145.37	147.94	142.10	134.18	148.86
14 hr	133.48	139.17	128.65	132.27	122.32	117.33	126.75	127.02	123.51	135.72
15 hr	123.52	139.90	137.85	137.93	129.92	152.93	138.05	157.19	160.41	142.59
16 hr	128.97	126.62	141.51	128.63	129.64	131.02	125.17	133.45	120.94	131.26
17 hr	117.95	112.12	112.13	142.75	117.29	109.58	135.28	146.79	120.97	122.53
18 hr	142.46	150.45	121.81	128.21	126.94	118.43	114.89	145.34	126.71	115.92
19 hr	122.32	116.80	107.37	124.90	132.70	135.90	118.05	118.20	111.46	106.27
20 hr	111.71	128.00	115.01	100.59	107.21	118.75	104.11	99.36	99.69	102.85
21 hr	101.99	112.60	130.15	125.07	128.54	115.37	132.58	121.26	110.80	129.38
22 hr	140.97	124.73	118.20	143.65	132.67	135.25	121.33	110.09	122.31	119.30
23 hr	106.59	104.11	107.61	100.60	95.22	96.84	92.13	90.22	100.87	98.68

REFERENCES

Akaike, H. (1969) Fitting autoregressive models for prediction, *Ann. Inst. Statist. Math.*, **21**, 243–247.

Akaike, H. (1977) On entropy maximisation principle, In *Applications of Statistics*, ed. P.R. Krishnaiah, 27–41, New Holland: Amsterdam.

Arató, M. (1982) *Linear stochastic systems with constant coefficients: a statistical approach*, Lecture notes in control and information sciences, **45**, Springer-Verlag: Berlin.

Ash, R.B. and Gardner, M.F. (1975) *Topics in stochastic processes*, Academic Press: New York.

Atkinson, J.D. (1967) *Spectral density of first order piecewise linear systems excited by white noise*, Ph.D. thesis, California Institute of Technology.

Atkinson, J.D. and Caughey, T.K. (1968) Spectral density of piecewise linear first order systems excited by white noise, *J. Non-Linear Mechanics*, **3**, 137–156.

Bartlett, M.S. (1946) On the theoretical specification and sampling properties of autocorrelated time-series, *J. Roy. Statist. Soc. Ser. B*, **8**, 27–41.

Bartlett, M.S. (1955) *An introduction to stochastic processes with special reference to methods and applications*, Cambridge University Press: London.

Berger, J.O. (1985) *Statistical decision theory and Bayesian analysis,* second edition, Springer-Verlag: New York.

Bergstrom, A.R. (1984) Continuous time stochastic models and issues of aggregation over time, In *Handbook of econometrics*, Vol. 2, ed. Griliches, Z. and Intriligator, M.D., Elsevier Science Publishers: BV, 1146–1212.

Billingsley, P. (1968) *Convergence of probability measures*, John Wiley & Sons: New York.

Box, G.E.P. and Jenkins, G.M. (1976) *Time series analysis: forecasting and control,* revised edition, Holden-Day: San Francisco, California.

Box, G.E.P. and Tiao, G.C. (1973) *Bayesian inference in statistical analysis*, Addison-Wesley: Reading, Massachusetts.

Breiman, L. (1968) *Probability*, Addison-Wesley: Reading, Massachusetts.

Brent, R.P. (1973) *Algorithms for minimization without derivatives,* chapters 3–4, Prentice-Hall: Englewood Cliffs, N.J..

Brockwell, P.J. and Davis, R.A. (1991) *Time series: theory and methods,* second edition, Springer-Verlag: New York.

Brockwell, P.J. and Hyndman, R.J. (1992) On continuous-time threshold autoregression, *Int. J. Forecasting*, **8**, 157–173.

Brockwell, P.J., Hyndman, R.J. and Grunwald, G.K. (1991) Continuous time threshold autoregressive models, *Statistica Sinica*, **1**, No. 2, 401–410.

Brockwell, P.J., Resnick, S.I. and Pacheco-Santiago, N. (1982) Extreme values, range and weak convergence of integrals of Markov chains, *J. Appl. Prob.*, **19**, 272–288.

Chan, K.S., Petruccelli, J.D., Tong, H. and Woolford, S.W. (1985) A multiple threshold AR(1) model, *J. Appl. Prob.*, **22**, 267–279.

Chan, K.S. and Tong, H. (1986) On estimating thresholds in autoregressive models, *J. Time Series Anal.*, **7**, 179–190.

Chan, K.S. and Tong, H. (1987) A note on embedding a discrete parameter ARMA model in a continuous parameter ARMA model, *J. Time Series Anal.*, **8**, 277–281.

Chen, R. and Tsay, R.S. (1991) On the ergodicity of TAR(1) processes, *Ann. Appl. Prob.*, **1**, 613–634.

Conway, E.D. (1971) Stochastic equations with discontinuous drift, *Trans. Amer. Math. Soc.*, **157**, 235–245.

Conway, E.D. (1972) Stochastic equations with discontinuous drift, II, *Indiana Univ. Math. J.*, **22**, 91–99.

Cox, D.R. and Miller, H.D. (1965) *The theory of stochastic processes*, Methuen & Co: London.

Cramér, H. and Leadbetter, M.R. (1967) *Stationary and related stochastic processes: sample function properties and their applications*, John Wiley and Sons: New York.

Davies, N., Pemberton, J. and Petruccelli, J.D. (1988) An automatic procedure for identification, estimation and forecasting univariate self exciting threshold autoregressive models, *The Statistician*, **37**, 199–204.

Doob, J.L. (1953) *Stochastic processes*, John Wiley and Sons: New York.

Durbin, J. (1961) Efficient fitting of linear models for continuous stationary time series from discrete data, *Bull. Int. Statist. Inst.*, **38**, 273–281.

Dzhaparidze, K.O. (1971) On methods for obtaining asymptotically efficient spectral parameter estimates for a stationary Gaussian process with rational spectral density, *Theory Prob. Appl.*, **16**, 550–554.

Gantmacher, F.R. (1960) *The Theory of Matrices,* Vol. 2, Chelsea Publishing Company: New York.

Gerald, C.F. and Wheatley, P.O. (1989) *Applied Numerical Analysis,* Fourth Edition, Addison-Wesley: Reading, Massachusetts.

Ghaddar, D.K. and Tong, H. (1981) Data transformation and self-exciting threshold autoregression, *Appl. Statist.*, **30**, 238–248.

Gikhman, I.I. and Skorohod, A.V. (1969) *Introduction to the theory of random processes*, W.B. Saunders Company: Philadelphia.

He, S.W. and Wang, J.G. (1989) On embedding a discrete-parameter ARMA model in a continuous-parameter ARMA model, *J. Time Series Anal.*, **10**, 315–323.

Henrici, P. (1962) *Discrete variable methods in ordinary differential equations*, John Wiley & Sons: New York.

Hurwitz, A. (1895) Ūber die Bedingungen, unter welchen eine Gleichung nur Wurzeln mit negativen reellen Teilen besitzt, *Math. Ann.*, **46**, 273–284.

Hyndman, R.J. (1990) *Constructing highest density regions*, Research Report No. 4, Dept. of Statistics, University of Melbourne, May 1990.

Hyndman, R.J. (1992) Yule–Walker Estimates for Continuous Time Autoregressive Models, *J. Time Series Anal.*, forthcoming.

Jazwinski, A.H. (1970) *Stochastic processes and filtering theory*, Academic Press: New York.

Jones, R.H. (1981) Fitting a continuous time autoregression to discrete data, In *Applied Time Series Analysis II*, ed. D.F. Findlay, Academic Press: New York, 651–682.

Jones, R.H. (1984) Fitting multivariate models to unequally spaced data, In *Time series analysis of irregularly observed data*, Lecture notes in statistics, **25**, Springer–Verlag: New York, 158–188.

Kaplan, W. (1962) *Operational methods for linear systems*, Addison-Wesley: Reading, Massachusetts.

Karlin, S, and Taylor, H.M. (1981) *A second course in stochastic processes*, Academic Press: New York.

Khazen, E.M. (1961) Estimating the density of the probability distribution for random processes in systems with non-linearities of piecewise-linear type, *Theory Prob. Appl.*, **6**, 214–220.

Kurtz, T.G. (1975) Semigroups of conditioned shifts and approximation of Markov processes, *Ann. Prob.*, **3**, 618–642.

Lebedev, V.A. (1976) On a condition for uniqueness of the solution of a system of stochastic differential equations, *Theory Prob. Appl.*, **21**, 412–419.

Lebedev, V.A. (1978) On the uniqueness of a relaxed solution of a system of stochastic differential equations, *Theory Prob. Appl.*, **23**, 147–155.

Lee, P.M. (1989) *Bayesian statistics: an introduction*, Oxford University Press: New York.

Liénard and Chipart (1914) Sur la signe de la partie réelle des racines d'une équation algébrique, *J. Math. Pures Appl. (6)*, **10**, 291–346.

Milshtein, G.N. (1974) Approximate integration of stochastic differential equations, *Theory Prob. Appl.*, **19**, 3, 557–562.

Milshtein, G.N. (1978) A method of second-order accuracy integration of stochastic differential equations, *Theory Prob. Appl.*, **23**, 2, 396–401.

Moler, C. and Van Loan, C. (1978) Nineteen dubious ways to compute the exponential of a matrix, *SIAM Rev.*, **20**, 801–836.

Moran, P.A.P. (1953) The statistical analysis of the Canadian lynx cycle, I: structure and prediction, *Aust. J. Zool.*, **1**, 163–173.

Øksendal, B. (1989) *Stochastic differential equations: an introduction with applications*, second edition, Springer–Verlag: Berlin.

Ozaki, T. (1985) Non-linear time series models and dynamical systems, In *Handbook of statistics*, Vol. 5, ed. E.J. Hannan, P.R. Krishnaiah, M.M. Rao, pp.25–83. Elsevier Science Publishers, B.V..

Pandit, S.M. and Wu, S.M. (1983) *Time series and system analysis with applications*, John Wiley and Sons: New York.

Pemberton, J. (1989) Forecasting accuracy of non-linear time series models., Tech. Rep., Worcester Polytechnic Institute, MA.

Petruccelli, J.D. (1990) A comparison of tests for SETAR-type non-linearity in time series, *J. Forecasting*, **9**, 25–36.

Phadke, M.S. and Wu, S.M. (1974) Modelling of continuous stochastic processes from discrete observations with applications to sunspot data, *J. Amer. Statist. Assoc.*, **69**, 325–329.

PHILLIPS, A.W. (1959) The estimation of parameters in systems of stochastic differential equations, *Biometrika*, **46**, 67–76.

PHILLIPS, P.C.B. (1972) The structural estimation of a stochastic differential equation system, *Econometrica*, **40**, 1021–1041.

PLATEN, E. (1987) Derivative free numerical methods for stochastic differential equations, In *Stochastic differential systems: proceedings of the IFIP-WG 7/1 working conference, Eisenach, GDR, April 6–13, 1986.* Ed. H.J. Engelbert and W. Schmidt, Lecture notes in control and information sciences, **96**, Springer–Verlag: Berlin.

PRABHU, N.U. (1965) *Stochastic processes*, Macmillan: New York.

PRESS, W.H., FLANNERY, B.P., TEUKOLSKY, S.A. and VETTERLING, W.T. (1988) *Numerical recipes in C: the art of scientific computing*, Cambridge University Press: Cambridge.

PRIESTLEY, M.B. (1981) *Spectral analysis and time series, Vol. 1: univariate series*, Academic Press: London.

ROBINSON, P.M. (1977) Estimation of a time series model from unequally spaced data, *Stoch. Proc. Appl.*, **6**, 9–24.

ROBINSON, P.M. (1980) Continuous model fitting from discrete data, In *Directions in time series*, ed. Brillinger, D. and Tiao, G., 263–278.

ROUTH, E.J. (1877) *A treatise on the stability of a given state of motion*, Macmillan: London.

ROUTH, E.J. (1905) *The advanced part of a treatise on the dynamics of a rigid body*, sixth edition, Macmillan: London; repr. 1959, Dover: New York.

STROOCK, D.W. and VARADHAN, S.R.S. (1969) *Diffusion processes with continuous coefficients I, II*, Comm. Pure Applied Math., **22**, 345–400, 479–530.

TONG, H (1978) On a threshold model, In *Pattern recognition and signal processing*, ed. C H Chen, Sijhoff and Noordhoff: Amsterdam, 575–586.

TONG, H (1983) *Threshold models in non-linear time series analysis*, Lecture notes in statistics, **21**, Springer-Verlag: New York.

TONG, H. (1989) Thresholds, stability, nonlinear forecasting and irregularly sampled data, In *Statistical analysis and forecasting of economic structural change*, ed. P. Hackl, pp.279–296, Springer–Verlag: Heidelberg.

TONG, H. (1990) *Non-linear time series: a dynamical system approach*, Oxford University Press: Oxford.

TONG, H. and LIM, K.S. (1980) Threshold autoregression, limit cycles and cyclic data (with discussion), *J. Roy. Stat. Soc. Ser. B*, **42**, 245–292.

TONG, H. and MOEANADDIN, R. (1988) On multi-step non-linear least squares prediction, *The Statistician*, **37**, 101–110.

TONG, H. and PEMBERTON, J. (1980) On the stability and limit cycles of non-linear autoregression in discrete time, *Cahiers du CERO*, **22**, 2, 137–148.

TONG, H. and YEUNG, I. (1990) On tests for threshold-type nonlinearity in irregularly spaced time series, *J. Statist. Comput. Simul.*, **34**, 177–194.

TONG, H. and YEUNG, I. (1991) Continuous time threshold autoregressive models, *Statistica Sinica*, **1**, 411–430.

TSAY, R.S. (1989) Testing and modeling threshold autoregressive processes, *J. Amer. Statist. Assoc.*, **84**, 231–240.

WENTZELL, A.D. (1981) *A course in the theory of stochastic processes*, McGraw-Hill: New York.

WRIGHT, D.E. (1986) A note on the construction of highest posterior density intervals, *Appl. Statist.*, **35**, 49–53.

YAGLOM, A.M. (1987) *Correlation theory of stationary and related random functions, Vol. 1: basic results*, Springer-Verlag: New York.

YEUNG, I. (1989) *Continuous time threshold autoregressive models*, Unpublished Ph.D. thesis, University of Kent.

YULE, G.U. (1927) On a method of investigating periodicities for Wolfer's sunspot data, *Phil. Trans. Roy. Soc., Series A*, **226**, 267–298.

ZVONKIN, A.K. (1974) Transformation of the phase space of a diffusion process eliminating drift, *Matem, S.*, **93** (135), 129–149. (In Russian. Quoted by Lebedev, 1976).